JISC Shared Infrastructure Services Synthesis Study

A review of the shared infrastructure for the JISC Information Environment

Document details

Author:	Ann Chapman and Rosemary Russell
Date:	28 September 2006
Version:	Final
Document Name:	jisc-sis-report-final.doc
Notes:	

Acknowledgements

UKOLN is funded by the MLA: The Museums, Libraries and Archives Council, the Joint Information Systems Committee (JISC) of the Higher and Further Education Funding Councils, as well as by project funding from the JISC and the European Union. UKOLN also receives support from the University of Bath where it is based.

Contents Page

JISC Shared Infrastructure Services Synthesis Study

A review of the shared infrastructure for the JISC Information Environment

1 Executive Summary

The JISC aims to provide a first-class sustainable infrastructure using a common, integrated information and communications environment for HE/FE. JISC programmes have funded the development of a number of component parts, though many are still in the development pipeline and some gaps remain. Some of these gaps may be filled by products and services developed outside the JISC community or in collaboration.

This small-scale study has reviewed the component parts identified in the JISC IE architecture, the projects both within and outside the JISC community that are developing relevant tools and services, their current stage of development and the risk to the JISC IE should they not proceed to full service. A focus group workshop has identified a number of key issues that concern the community and that the JISC needs to address.

Overall shared infrastructure services components are still largely in development. The provision of a testbed environment will be helpful but pilots and trials will still be needed and some services need to be populated with data. Although some of this needs to take place within the JISC community, it is also an area where there is potential benefit for collaboration with other sectors and JISC should actively encourage and support such partnerships.

The review has identified some areas where new projects and/or studies are required; these are noted in the recommendations. In some cases, the recommendations propose collaboration with partners outside the JISC community, and section 12 reports on some potential collaborative areas.

It is helpful that the JISC is now developing a project development lifecycle, with a projected pathway from exploratory project through to full service delivery (although not all projects will progress past the exploratory stage). However, as projects reach the point of business case modelling they are currently not given JISC guidelines on the type of metrics required or methods of establishing cost-benefit, including intangible benefits. Additionally, when they approach potential fee-paying users of their service, the interest shown is dependent on the service having supported long-term viability.

There is a need for the JISC to promote shared infrastructure services to the community, particularly once a service comes on stream. This will require the production of sufficient and appropriate documentation (including installation guides) for the service, and promotional materials (perhaps drawing on the scenarios from this report) to encourage take up by institutions.

The JISC also needs to demonstrate ‘best practice’ by providing more specific guidance to software developers to ensure the creation of reliable, good quality products. Best practice should also be encouraged in the deposit and long-term preservation of software. This might, for example, build on approaches like Open Middleware Infrastructure Institute (OMII) or be a research issue for JISC preservation programmes.

Both within the JISC community and without there are richer and poorer institutions in terms of technical ability and funding to support buy-in to shared infrastructure services. The move to a business model should not create a two-tier situation, and where possible methods of supporting the poorer institutions should be sought.

Digital policy management is a key factor in enabling users to access a wide range of resources. Provision of shared infrastructure services in this arena will best be developed in collaboration with content providers and to international standards for licensing terms

messages. Authentication and authorisation will also be required but is out of scope for this report; it is noted that the launch of the Access Management Federation in September 2006 will provide new opportunities in this area.

JISC needs to make sure that the approach to project lifecycles and shared infrastructure service progress is made clear to the projects working in this area and the wider community. We note that this report will go someway towards that as will the detailed work that JISC is undertaking on project lifecycles.

1.1 Recommendations

The following recommendations have been made.

1.1.1 The JISC Information Environment and Resource Discovery.

- ***Recommendation: shared infrastructure services need to be extremely reliable (since failure of a shared service will result in failure or problems at multiple end-user services downstream). Therefore, all shared infrastructure services should be hosted at their own DNS domain name, replicated on at least two servers at national data-centres, using DNS-hiding to hide multiple servers from client services.***
- ***Recommendation: need discussion with community about 'trust' issues with shared infrastructure services; this could include reliability, quality of service, integrity of content, privacy.***
- ***Recommendation: most shared infrastructure services should be designed with multiple instances in mind, i.e. that there will be multiple different instances of any given shared service (as opposed to replicating the same service at multiple points on the network). Cooperation between such instances should be encouraged on a national and international basis - e.g. sharing of underlying data, etc. (An example would be the IEMSR, the DC registry and a European eLearning metadata schema registry which may want to share data.)***
- ***Recommendation: shared infrastructure services built on aggregations of data (service registries, metadata schema registries, terminology services, etc.) should encourage data owners to publish the underlying data publicly on the Web, using appropriate RDF/OWL and/or XML schemas and assigning persistent 'http' URIs to the key entities within the data. Such services should primarily see themselves as authoritative aggregators of publicly available data, rather than as the master copy of that data.***
- ***Recommendation: machine interfaces to shared infrastructure services should be driven by community need and underlying functional requirements but should in general be as lightweight as possible.***
- ***Recommendation: as far as possible, machine interfaces to shared infrastructure services should adhere to widely adopted international standards (formal or de facto). The JISC community should work towards reaching global agreements on such interfaces wherever possible.***
- ***Recommendation: although human-interfaces to shared infrastructure services are necessary, the developers of shared services should focus on the machine interface. Whenever possible, the human interface to a shared infrastructure service should make use of the underlying machine interface.***
- ***Recommendation: The funding, development and deployment of 'common services' (i.e. services that are useful across a number of domains of use) should be coordinated across the digital library, e-learning, e-research and e-administration domains; this will be helped by the JISC e-Framework approach***

1.1.2 Digital Policy Management

- ***Recommendation: JISC should develop an over-arching architectural strategy for digital policy management.***

- ***Recommendation: JISC should undertake a formal gap analysis to define areas in which JISC itself needs to act as a priority and those where it is more appropriate to wait for developments elsewhere.***

1.1.3 Personalisation

- ***Recommendation: JISC should consider looking at whether shared infrastructure services could/should provide additional support for personalisation in order to enable more adaptive personalisation within the IE.***

1.1.4 Identifier Services

- ***Recommendation: an OAI-PMH interface for the OpenURL Router would be useful, allowing OpenURL source services to harvest the data to part-populate their own internal registries.***
- ***Recommendation: JISC/OCLC/Digital Library Federation (DLF) should cooperate on a global solution to the problem of seamlessly discovering the correct OpenURL resolver.***
- ***Recommendation: JISC should consider the provision of a default OpenURL resolver for smaller/poorer institutions, in collaboration with a partner such as OCLC Openly Informatics or EDINA (possibly based on ZBLSA work).***

1.1.5 Representation Information and Format Registries

- ***Recommendation: At some point before the end of DCC funding, JISC should review the RI RegRep for potential long-term support.***
- ***Recommendation: JISC should talk to The National Archives (TNA) re next steps for PRONOM. There may be potential for TNA to collaborate with other projects.***
- ***Recommendation: JISC support the potential for format conversion work within PRONOM and DCC work rather than seek to develop any new service.***
- ***Recommendation: JISC should keep in touch with Global Digital Format Registry (GDFR) progress.***

1.1.6 Managing Digital Resources

- ***Recommendation: Carry out a small-scale study to examine synergies between ROMEO and ONIX for Licensing Terms.***
- ***Recommendation: Carry out an evaluation for funding ROMEO development to M2M capacity.***
- ***Recommendation: JISC should maintain contact with BIC and seek collaboration opportunities (tool building and piloting) in the continuing development of ONIX for Licensing Terms.***
- ***Recommendation: JISC should investigate the need for a licence registry shared service component, initially by funding a pilot project.***

1.1.7 Service Registries and supporting metadata

- ***Recommendation: it would be useful for JISC to offer some guidance on IESR content scope, in terms of coverage (eg JISC IE, UK-wide, European, international), and resource type (repositories etc).***
- ***Recommendation: resourcing for populating the IESR database should be considered.***
- ***Recommendation: further discussion is needed about the issue of distributed IESR registries.***

- ***Recommendation: collaboration should be pursued with OpenDOAR in order to ensure compatibility with IESR.***
- ***Recommendation: a realistic 'business plan' should be developed for IESR; it is unlikely that service registries can be sustainable without some form of support.***
- ***Recommendation: JISC should continue to fund work to develop and maintain a standard for collection-level description metadata.***

1.1.8 Metadata Schema Registries

- ***Recommendation: JISC may need to consider whether separate services are required to manage DC, LOM (and other future) schemas.***
- ***Recommendation: JISC should consider how to encourage the IEMSR registry to be populated in order to achieve a critical mass of data.***
- ***Recommendation: A 'collection policy' for inclusion of schemas needs to be agreed.***

1.1.9 Institutional profiling services

- ***Recommendation: JISC should review the approach recommended by the EDINA study and discuss practical implementation with IESR and other relevant projects.***

1.1.10 Terminology Services

- ***Recommendation: JISC should come to a decision on whether GeoCrossWalk can deliver; if yes, approve transition to service.***
- ***Recommendation: JISC should fund a technical review of GeoCrossWalk as part of the transition to service.***
- ***Recommendation: JISC should consider funding a discrete section of the IE to be fully operational (e.g. by populating databases) in order to demonstrate full functionality; this could potentially build on the IE Testbed approach.***
- ***Recommendation: HILT should undertake in-depth user testing in the context of significant (but contained and context-specific) mapping work.***
- ***Recommendation: It would be useful for HILT and the Becta Vocabulary Studio to explore collaborative work, especially given HILT's plans to examine the possibilities of a distributed approach to terminology services and inter-scheme mapping.***
- ***Recommendation: JISC should investigate further the potential utility of HILT and the BECTa service to the repositories programme.***
- ***Recommendation: JISC should consider alternative approaches to terminology services, including ontologies, text mining and folksonomies.***

1.1.11 Name Authority

- ***Recommendation: name authority for repositories requires further investigation, including the option of HESA identifiers.***
- ***Recommendation: the option of developing a UK name authority should be investigated.***
- ***Recommendation: JISC should work with other interested bodies including the British Library, and consider harnessing the enthusiasm of The National Archives (TNA) to lead a collaborative UK name authority effort.***
- ***Recommendation: collaboration with the SURF DAREnet name authority initiative should be explored.***

1.1.12 Working relationships

- ***Recommendation: JISC and shared infrastructure services need to develop a common understanding of the project lifecycle, so that stakeholders can be assured of continuity within the limitations of short term funding cycles imposed on government agencies such as JISC.***
- ***Recommendation: JISC need to promote a better understanding of the IE in the wider community, so that larger vendors are more prepared to work with JISC services.***
- ***Recommendation: JISC should maintain regular contact and consultation with other key organisations, both in the UK and internationally, in order to facilitate early identification of potential collaboration and synergy.***
- ***Recommendation: JISC should work with other key organisations to ensure interoperability between any e-infrastructure components that support Grid and e-Research and the IE. The IE itself should be developed with such interoperability in mind.***

1.1.13 Software Quality Assurance and Testing Management

- ***Recommendation: JISC should fund a small background study prior to commissioning (or inviting to tender) a software quality assurance and testing management system (SQATM) package.***
- ***Recommendation: JISC should produce guidelines regarding the deposit and long-term preservation of software in appropriate locations.***
- ***Recommendation: Commissioning of any (pilot or production) should include, in addition to working software) a complete package of materials (e.g. the 'how to' guide for installation, glossy brochures for promotion, phone support, support/users e-list) to promote and facilitate institutional adoption.***

2 Introduction and Terms of Reference

One of the strategic aims of the JISC Strategy 2004-2006 is: "To develop solutions that enable the United Kingdom education and research communities to keep their activities world class through the innovative use of ICT – by providing a first-class sustainable infrastructure." A key priority within this strategic aim is "to develop a common, integrated information and communications environment". Shared infrastructure services are essential building blocks for an efficient and effective information and communications environment.[1]

The JISC programme thus far has been called Shared services but recently JISC has changed the name to Shared Infrastructure Services in order to make it clearer that the work is about services that operate as underlying machine-to-machine (M2M) shared services.

JISC, having funded a variety of projects over the last few years that are component parts of the JISC IE, now needs to looks at what is missing from the jigsaw, and devise a strategy to fill the gaps. In addition to these programmes of work, since JISC works in a mixed economy environment where some services, tools and content are available from other sectors, it is recognised that some gaps may be best – or more appropriately – filled by products and services developed in other UK sectors, or internationally or commercial.

JISC has commissioned UKOLN to undertake a small-scale study that will develop an overview specification for shared infrastructure; the resulting direction-setting document is to be used to support the JISC Call for Projects planned for September 2006.

3 Aims and Objectives of the study

The aims are:

- To inform development of machine to machine (M2M) shared infrastructure for resource discovery, digital rights management, repositories and preservation;
- To synthesise requirements across effort so far;
- To help stakeholders understand the need for this M2M shared infrastructure.

The objectives are:

- To develop a set of scenarios that demonstrate how and why shared infrastructure is required;
- To identify requirements from documentation from JISC projects and foundation architecture papers;
- To synthesise the outcomes of effort to date, from JISC activities and the wider context;
- To include where possible the international and commercial context;
- To identify risks, additional to those already identified by JISC, in the area of shared infrastructure services: the current study (in particular the scenarios and requirements synthesis) will help to mitigate the risks; recommendations for further mitigation will be made where relevant;
- To provide a report that transforms the study findings into a series of direction-setting recommendations.

4 Methodology

In order to achieve the aims and objectives, the study focused on identifying relevant information in project documentation, supplemented by information obtained through email, telephone calls and a small number of face-to-face meetings. It also used expertise available at UKOLN and consultancy services in specialist areas.

- The study strands were to:
    - Develop a set of scenarios that will illustrate the need for M2M shared infrastructure and potential use
    - Identify objectives and deliverables, current status and potential interaction with other services for shared infrastructure services projects
    - Obtain views from the wider community on key issues
    - Incorporate feedback into final report
- Scope and boundaries of the work
    - Consultation with JISC shared infrastructure services and programmes and with other relevant organisations
    - Because of the short timescale, the study concentrated on the most relevant key contacts
    - The study covered the requirements of new areas that the shared infrastructure services need to support – repositories and digital preservation in particular
    - An international perspective has been sought via email and possible telephone interview using UKOLN contacts.
- Consultancy services were used in three key areas for their expertise:
    - Rightscom was commissioned to provide a synthesis of existing digital rights management work and a commentary on the maturity of available technologies, commercial issues and requirements for future development
    - Eduserv was commissioned to cover of licensing issues and a synthesis of existing work on architecture and shared infrastructure

- Leona Carpenter was commissioned to contribute to the study, principally by detailing the scenarios and facilitating the feedback workshop.

Note: Authentication and authorisation are outside the scope of this study.

5 The JISC Information Environment and resource discovery

Contribution by Andy Powell, Eduserv

The JISC Information Environment (JISC IE)[2] technical architecture specifies a set of standards and protocols that support the development of an integrated set of networked services for use by the UK HE and FE community. The intention is to allow the end-user to more seamlessly discover, access, use and publish digital and physical resources as part of their learning and research activities.

The key standards and protocols specified in the technical architecture are listed in the JISC IE Technical Standards[3].

In the JISC IE, a 'service component' is a network service, i.e. a service that is provided on-line. Example network services include Web sites, document supply services, abstracting and indexing services, data archives, online catalogues, databases, email archives, format conversion services, printing services, authentication or e-commerce services, etc. Each service component may offer part or all of its functionality through one or more Web services.

The kinds of service components made available through the JISC IE are shown in the diagram below. This diagram is not intended to be definitive. However, it is worth noting that at the time of writing, the majority of these components have been instantiated in some form or other as real service components on the network [4].

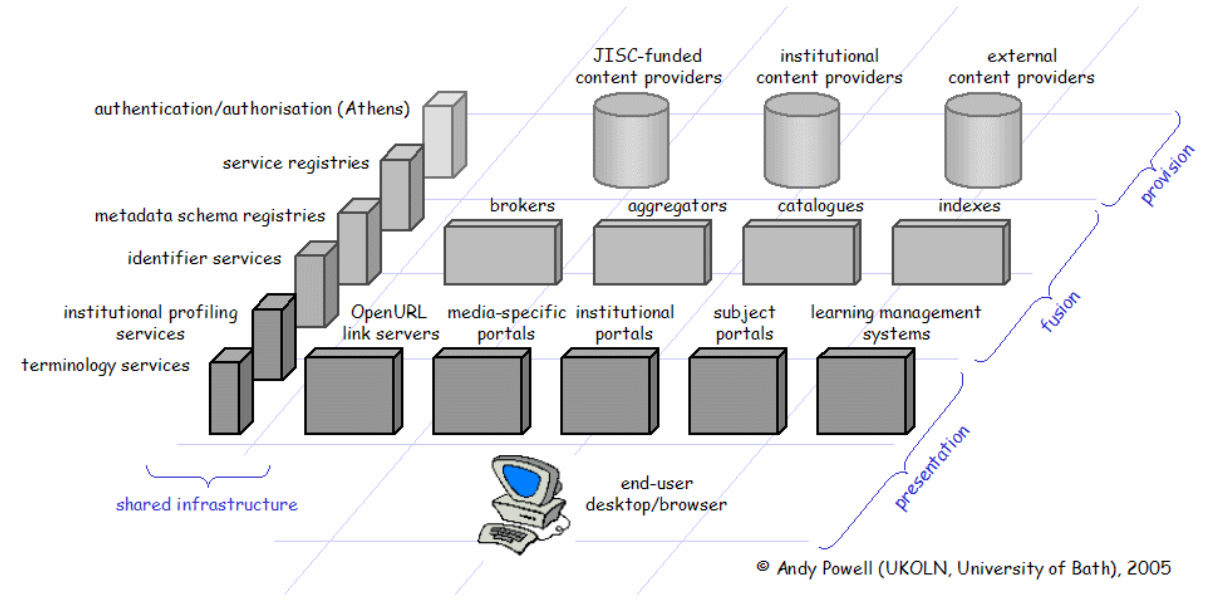

Figure 1 - The JISC IE architecture diagram

As can be seen from the diagram above, the JISC IE architecture envisages a number of shared infrastructure service components. Infrastructural services are defined as "a range of … network services that are called on by content providers, brokers, aggregators, indexes, catalogues and portals. Infrastructural services include authentication, authorisation, service registry, user preferences, resolver, institutional profile, metadata schema registry and terminology services".

At the time that the JISC IE architecture diagram was developed a number of such services were envisaged, including:

Authentication/authorisation Service

A structured network service that determines that the digital ID being presented to a network service is being used by the real-world individual who has the rights to use it and whether a particular digital ID has the necessary access-rights to access a particular resource. This is often achieved through the use of a username/password combination or a digital certificate, depending on the degree of assurance required.

Service Registry

A network service that stores and makes available descriptions of (i.e. metadata about) services and the content of collections made available through those services. A service registry is used by portals to determine what collections are available to end-users, and by portals, brokers and aggregators to determine how to interact with available network services.

Identifier Service

A network service that maintains and provides an association between an identifier and some metadata about the identified resource. Typically, an identifier service takes an identifier of a resource and returns a locator for it (usually in the form of a URL).

Institutional Profiling Service

A structured network service that stores and makes available information about what licences institutions hold, i.e. their access rights as organisations to particular resources, and other institution-wide preferences, such as preferred content-delivery services.

Metadata Schema Registry

A network service that stores and makes available information about the metadata schemas in use by other services.

Terminology Service

A structured network service that offers terminology-related services, for example mapping a term from one controlled vocabulary to another or expanding terms within a thesaurus.

It should be noted that none of these candidate service components is intended to be deployed on a single-instance basis – i.e. there are likely to be more than one of each of these service components deployed within (and beyond) the UK HE and FE community. Furthermore, there will be a range of additional shared infrastructural services supporting other functional requirements such as preservation (format registries, format conversion services, etc.) and e-learning (content packaging services, group management, etc.).

As a concrete example of this, it is worth noting that the OpenURL Router[5] service component has been developed and deployed since the original conception of the diagram. The OpenURL Router is a service component that routes requests for OpenURL link resolver services back to the correct institutional OpenURL resolver. (See also section 11.1.4.)

In passing, it is perhaps also worth noting that the treatment of authentication/authorisation in the JISC IE architecture diagram is somewhat simplistic, particularly in the context of the transition by the community from Athens to Shibboleth[6]. In a Shibboleth environment there is no authorisation service component as such. Authorisation is determined by the other service components based on the various attributes passed to them by the Shibboleth Identity Provider (IdP) service component. The Identity Provider components can be thought of as part of the shared infrastructure of the JISC IE in the sense that they are called on by many of the other service components on the diagram, but they will typically be deployed within each institution. The Shibboleth Service Provider functionality will be wrapped into the human-oriented user-interfaces of the other service components. Finally, the Where are you From (WAYF) service will need to be deployed as a shared infrastructural service component, routing requests back to the institutional Identity Provider components.

Given that there are some similarities in the nature of the OpenURL Router and the Shibboleth WAYF service, it may be sensible to consider co-locating these services on the network.

It should be noted that the range of personal attributes managed and exposed by the Identity Provider service component in a Shibboleth environment could be used to personalise the functionality of other service components.

JISC has deployed a development Service Registry component in the form of the JISC IE Service Registry (IESR) and has built up quite a lot of experience in the design of such a service. In meetings between the IESR project and other similar initiatives, such as the OCKHAM project funded by the NSF in the US, it has become clear that mechanisms for distributing the 'service registry' component across the network will be required (in much the same way that the DNS name lookup service is distributed across many network nodes).

Many identifier services are currently deployed, most outside the context of the JISC IE architecture. Prominent examples include the Handle System[7], the DOI[8] and PURLs[9].

Since the development of the original JISC IE architecture diagram, a revised version has been prepared to show how digital preservation service might fit into the environment.

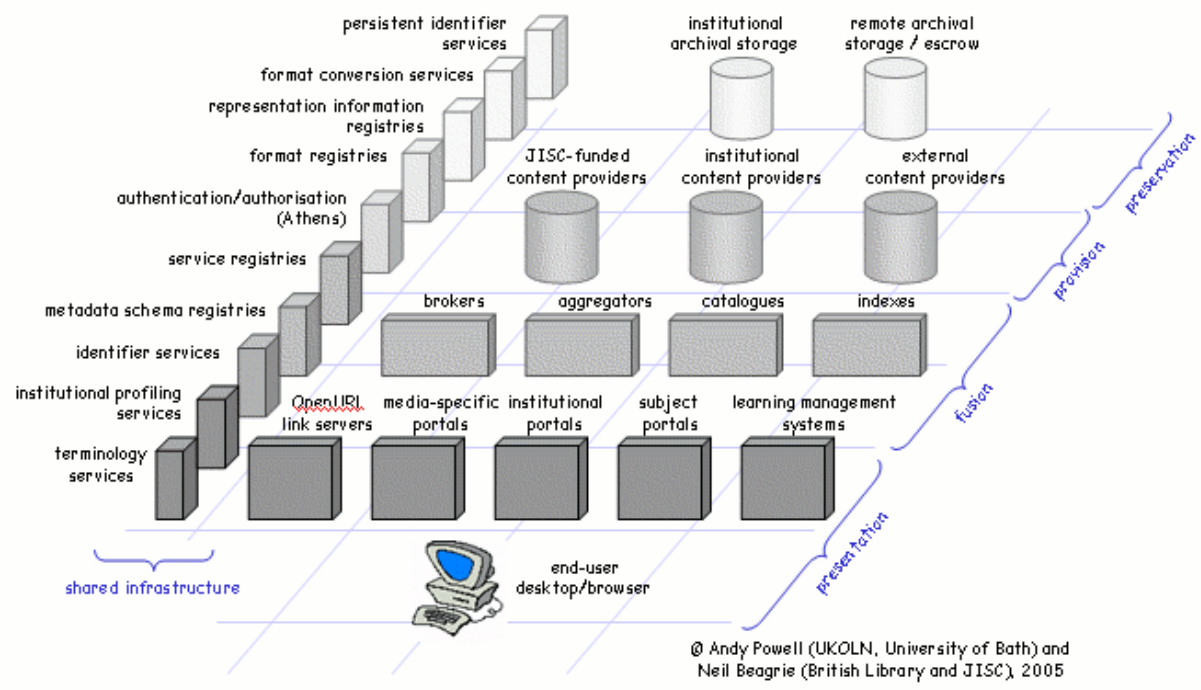

Figure 2 - The JISC IE and preservation

This diagram shows four additional classes of shared infrastructure services:

Representation Information Registry

A structured network service that maintains and exposes OAIS defined Representation Information (RI) associated with the digital objects in the JISC Information Environment. [Representation Information] (RI) is a term encompassing all information required to access the information stored within a digital object. The term can be applied to all levels of abstraction and refers to both the structural and semantic composition. The use of RI is often recursive: using one element of RI in a meaningful manner requires further RI. This recursion continues until the contents of the original object can be accessed and understood by the current designated community.

Format Registry

A structured network service that maintains and exposes information (metadata) about the various file formats (including different versions of those file formats) in use within the JISC Information Environment.

Note that Format Registry services offer a subset of the functionality offered by Representation Information Registry services.

Format Conversion Service

A structured network service that takes content in one format (for example RTF) and converts it to another format (for example PDF).

Note that format conversion services are highlighted here as a key preservation service, but there are many other types of services related to preservation (for example, emulation services). Broadly speaking, preservation services can be categorised into three groups; characterisation, preservation planning, and preservation action.

Persistent Identifier Service

An identifier service that is specifically designed to offer consistent services over very long periods of time.

5.1 The e-Framework for Education and Research

The e-Framework[10] is an initiative by JISC, the Department of Education, Science and Training (DEST) in Australia and partners to produce an evolving and sustainable, open standards based, service oriented technical framework to support the education and research communities. It builds on the e-Learning Framework and the JISC Information Environment, to support education, research and management information systems.

The e-Learning Framework applied a service oriented approach to unbundling the monolithic virtual learning environments (VLEs) that were typically being deployed by HE and FE institutions. The intention was to allow the functionality of a VLE to be delivered in an unbundled, component based form, using individual, loosely-coupled services in order to achieve a more flexible, standards based, extensible, future-proofed and ultimately better value platform on which to base e-learning services. However, it soon became clear that these arguments could similarly be applied to other areas of institutional ICT deployment, notably in the areas of e-research and e-administration. The e-Framework grew out of the community's desire to apply a service-oriented approach more broadly than just e-learning.

The e-Framework essentially takes a Web 2.0 approach to the delivery of services, though some might argue that any coordinated national activity necessarily implies a more rigid approach to the development of individual services than one might otherwise find in the Web 2.0 arena. Discussions are ongoing within the community about how best to model and describe services and work-flows and about how prescriptive the community should be in the adoption of specific technical approaches such as SOAP and REST.

A number of shared infrastructure services are likely to arise within this context, at institutional, regional, national and international levels. It is not yet clear exactly what this set of shared infrastructure services might be, though it is clear that there will be significant overlaps with the JISC IE services described above in the context of resource discovery and preservation. However, the broadly scoped nature of the e-Framework implies that the required shared infrastructure services will go significantly beyond those envisaged for the JISC IE. For example, in the context of e-learning there are likely to be shared services in the areas of re-usable learning objects, content packaging, e-portfolios, the automated generation of LOM, learning design engines, group management, collaborative tools and so on.

The e-learning services currently envisaged within the e-Learning Framework are as follows:

- Activity Management
- Assessment
- Competency
- Course Validation
- Curriculum
- Grading
- Learning Flow
- Marking
- Quality Assurance
- Reporting

- Resource List
- Sequencing
- Tracking
- ePortfolio

It is perhaps worth noting that the E-Learning Framework makes reference to 'common services' – the notion that some services will prove to be useful across more than one domain (e-learning, e-research, etc.). Whilst common services may be deployed as shared (infrastructural) services, it does not mean that this will always be the case.

5.2 JISC Future Strategy

- ***Recommendation: shared infrastructure services need to be extremely reliable (since failure of a shared service will result in failure or problems at multiple end-user services downstream). Therefore, all shared infrastructure services should be hosted at their own DNS domain name, replicated on at least two servers at national data-centres, using DNS-hiding to hide multiple servers from client services.***
- ***Recommendation: need discussion with community about 'trust' issues with shared infrastructure services; this could include reliability, quality of service, integrity of content, privacy.***
- ***Recommendation: most shared infrastructure services should be designed with multiple instances in mind, i.e. that there will be multiple different instances of any given shared service (as opposed to replicating the same service at multiple points on the network). Cooperation between such instances should be encouraged on a national and international basis - e.g. sharing of underlying data, etc. (An example would be the IEMSR, the DC registry and a European eLearning metadata schema registry which may want to share data.)***
- ***Recommendation: shared infrastructure services built on aggregations of data (service registries, metadata schema registries, terminology services, etc.) should encourage data owners to publish the underlying data publicly on the Web, using appropriate RDF/OWL and/or XML schemas and assigning persistent 'http' URIs to the key entities within the data. Such services should primarily see themselves as authoritative aggregators of publicly available data, rather than as the master copy of that data.***
- ***Recommendation: machine interfaces to shared infrastructure services should be driven by community need and underlying functional requirements but should in general be as lightweight as possible.***
- ***Recommendation: as far as possible, machine interfaces to shared infrastructure services should adhere to widely adopted international standards (formal or de facto). The JISC community should work towards reaching global agreements on such interfaces wherever possible.***
- ***Recommendation: although human-interfaces to shared infrastructure services are necessary, the developers of shared services should focus on the machine interface. Whenever possible, the human interface to a shared infrastructure service should make use of the underlying machine interface.***
- ***Recommendation: The funding, development and deployment of 'common services' (i.e. services that are useful across a number of domains of use) should be coordinated across the digital library, e-learning, e-research and e-administration domains; this will be helped by the JISC e-Framework approach***

6 Digital Policy Management: managing access to and use of digital resources

Contributed by Mark Bide, Rightscom

> *DRM in an academic environment should be an "enabler" not a "preventer". Its purpose is to let people work as freely as possible in the knowledge that they are both working within the bounds of the law of copyright and respecting the rights of others.*
>
> Duncan et al (2004) *Digital Rights Management: a study by Intrallect Ltd on behalf of JISC* [11]

For many users, the preferred solution for dealing with the complexity and uncertainty that IPRs bring to access and use policies in the network environment is simply to ignore intellectual property completely. The end to *any* attempt to manage access to and use of resources on the network would certainly bring about considerable simplification, but would have consequences both far-reaching and unpredictable.

At what might be described as the other end of the spectrum, some media businesses see simplification as implying a requirement to gather the complexity and uncertainty of rights under unitary control, with "all rights" in a creation being centred in a single organisation; access and use can then be managed through the application of technical protection measures, whether or not these respect the rights of users or the interests of creators.

If we are to find a middle way between these extremes, it seems likely to involve both some simplification of the environment within which policies are created, and the use of technology as an aid to implementation of those policies.

6.1 Digital Rights Management or Digital Policy Management?

The term "Digital Policy Management" (DPM) is to be preferred to "Digital Rights Management" (DRM), because of the baggage that the latter term carries with it. In many peoples' minds, DRM is directly equated with "technical protection measures" for the *enforcement* of compliance with policies, while we believe that the emphasis should be on the *expression* of policies. DPM is therefore used for preference throughout this contribution.

There is another reason why the use of "rights" may be misleading in this context. Not all access and use policies are dependent on intellectual property rights. Issues relating to privacy and commercial confidentiality may be equally significant in establishing policies, particularly in an institutional context.

6.2 Describing the uncertainty and complexity

Reports that have been submitted to JISC over the last two years have certainly served to stress the dual problem of uncertainty and complexity. These challenges persist throughout the lifecycle approach which has usefully been proposed by Intrallect (see Figure 3), although it is perhaps incomplete in taking no account of the recording and reporting processes which need to be included in any complete account of the management of access to and use of digital resources. The lifecycle terminology has been adopted for the purposes of this report.

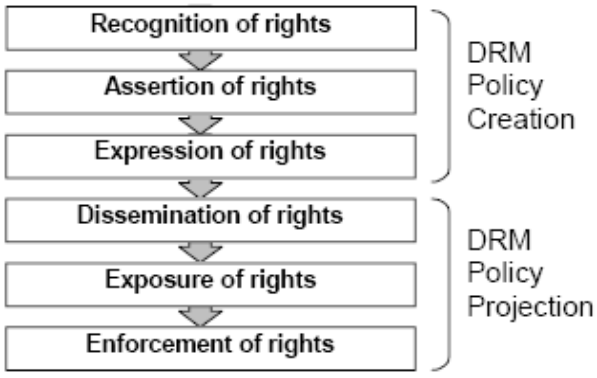

Figure 3: Intrallect's proposed "*Stages in the DRM Process*"

Much of the complexity and uncertainty relates to the first two elements of this lifecycle, which can be usefully paraphrased as:

- determining who owns the rights
- establishing mechanisms for the orderly transfer of rights and permissions through contract and licensing

These topics fall naturally outside the scope of this report, which is concerned only with the machine-to-machine (M2M) aspects of policy management; however, their resolution (or the failure to resolve them) will have a significant influence on what may be required in the technical environment.

6.3 Simplifying the rights and permissions environment

The thrust of the recommendations from the JISC *Rights in Digital Environments* workshops[12] held in March 2005 was to clarify and simplify the rights and permissions environment in FE and HE.

Presumably in response to this, the JISC IPR Consultancy[13] has (as one element of its work) an effort to clarify and standardise rights ownership within UK and FE institutions, and to standardise the terms on which digital creations are licensed within the educational community. There is a strong presumption in favour of "open" forms of trust-based licensing, under licences analogous to Creative Commons (CC),[a] to encourage the widest possible use of resources without commercial constraints.

A number of potential advantages can be identified for the FE/HE community from adopting this approach:

- A relatively undifferentiated policy environment (where users can be confident that all resources are available under one of a limited number of standard sets of terms of access and use) reduces the need to manage the complexity of the current highly differentiated licensing regime (and therefore reduces the cost of management)
- Adoption of a trust-based system, rather than one in which compliance is routinely enforced (whether technically or legally) further reduces cost, always assuming that real trust levels within the system can be kept high
- The motivation of creators of resources within the academic community may not be most effectively reflected by current commercial models of publication; a sustainable publication model which is better attuned to the real reward structures of the academic community should lead to the creation and dissemination of a greater volume of high quality resources, to the mutual advantage of both creators and users

Two recently commenced JISC projects, *TrustDR*[14] and *Rights and Rewards*[15] are exploring these themes, but are both at too early a stage for any results to be available.

However, regardless of the outcome of current projects, it seems unlikely to us that a simplified rights and permissions environment will be uniformly adopted for all digital resources required in the FE/HE sector, within any meaningful timeframe. There will continue to be a "mixed economy", in which commercially-motivated publication continues to play a significant role, and where the simplicity of an undifferentiated policy environment cannot be attained. While some reduction in the vagaries of current commercial licensing may be regarded as likely in the

[a] Jorum found that CC licences were unsuitable for its purposes, since they are essentially designed to support direct licensing from individual (creator) to individual (user). There is no capability to support other types of licensing arrangements which are required: from an institution to a service (like Jorum); or from the service to an end user. It is also important to recognise that Jorum is not strictly speaking an open access service; it is open only to registered users who are themselves members of the UK FE/HE teaching community using ATHENS-authenticated identities. Access is not available to other classes of user.

medium term, overall complexity in the licensing environment seems likely to be with us for a long time to come.[b]

It is in this context that we need to consider the role that technology may have to play in the management of rights for the FE/HE community.

6.4 The role of technology in the management of policy

If we return to the Intrallect model (Figure 3), we see technology as having a key role to play in the four remaining elements of the rights lifecycle which have been identified. Again we paraphrase:

- **expression of rights:** "languages" through which policies of access and use can be communicated (from simple tagging to complex XML expressions)
- **dissemination of rights:** mechanisms for associating the expressions with the resources to which they relate; this can involve simple tagging of the resource itself, or more complex repertoire management mechanisms
- **exposure of rights:** mechanisms by which the rights associated with specific resources can be communicated to human users (including, for example, using iconic languages such as that promoted by Creative Commons – or simply by presenting "click through" licences)
- **enforcement of rights:** technical protection measures (TPMs), through which compliance with policies can be assured where levels of trust are insufficient;[c] TPMs include not only those encryption technologies which we typically call "DRM" but also the access and authentication systems which we use every day

Additionally, we have already mentioned the requirement for **recording and reporting usage** which Intrallect did not include in their model; in the academic area, this has some obvious overlap with projects like Project COUNTER[16] and NISO's SUSHI.[17]

We therefore see the initial element of the technology challenge as essentially about the development and implementation of *metadata* and *identification* standards, rather than about application technology.

6.5 Moving forward

If the assumption can be made (as we believe that it can) that a heterogeneous policy environment for the management of digital resources in the FE/HE community is inevitable, and that simplification of that environment through administrative interventions can answer only part of the challenge, the development and implementation of an effective technology framework to manage policy appears to us to be essential.

There are a number of different technologies which will need to be integrated to create a solution, and these are not always seen as part of a coherent framework. The elements of the solution do not function in isolation from each other: there is little value in developing standards for the expression of policies in the absence of well-implemented disciplines for the unique identification of the resources to which those policies relate; the best designed framework for

---

[b] Consider the experience within even the open source community, where differences between different approaches to licensing have led to incompatibility and a requirement for interoperability – see the discussion on Lawrence Lessig's weblog (11 November 2005) http://creativecommons.org/weblog/entry/5709.

[c] The Intrallect model bundles forensic detection technologies, such as watermarking and fingerprinting, with enforcement technologies such as encryption. We would suggest that forensic technologies are probably more grouped with recording, reporting and auditing functions than with encryption. These technologies are, of course, most appropriate for use with non-text resources, including graphics, sound and audiovisual; we are not aware of any relevant work being undertaken by JISC, but nor are we aware of any requirements that have been expressed.

the communication of policies are of little value in the absence of the necessary technology for their interpretation, whether that interpretation is to inform a human user or an enforcement technology.

In the absence of an architecture and a prioritised gap analysis, we perceive the risks as lying in two dimensions, both linked to the lack of an evidence base:

- JISC might decide that no work is necessary on its part in the whole field of DPM
- JISC might continue to undertake work in the field of DPM outside a well-structured framework

And in the absence of a proper strategic framework, we believe that it will be impossible for JISC to respond appropriately to the comprehensive sets of recommendations that have been made.

- ***Recommendation: JISC should develop an over-arching architectural strategy for digital policy management.***
- ***Recommendation: JISC should undertake a formal gap analysis to define areas in which JISC itself needs to act as a priority and those where it is more appropriate to wait for developments elsewhere.***

7 Scenarios

By Leona Carpenter

The scenarios set out here are not intended to be exhaustive of the full range of activities that would be enabled by the implementation in the JISC Information Environment of shared infrastructure services elements described in this report. Rather, they are meant to provide an indicative overview of a range of these activities. They demonstrate a number of activities that would either be improved or made possible through the availability of this infrastructure. These scenarios focus on the actions (goal-directed behaviour) of imagined people as actors playing specific roles. A simple example of such an action is “search for a document that might be in a number of different repositories”. The goal to which this action is directed is "find the document as part of a literature search in the first step of a research project". The actor/role is "Sudeep, a research scientist".

To the extent possible, these scenarios draw on usage scenarios, use cases and functional requirements documents produced in Shared Infrastructure Services programme projects and requirements identified in programme studies. In addition to various presentations, these include: GeoCrossWalk Example Use Cases[18], HILT Phase III Use Cases, in M2M Pilot Demonstrator Project Proposal[19], IEMSR Phase 2 User Requirement documents[20], IE Services Registry project[21] and [22], Rights in Digital Environment report[23], Scoping study into Digital Rights Management[24] and the Scoping study into Institutional Profiling and Terms and Conditions Services[25].

In addition, it was useful to look back at the original JISC IE Architecture Usage Scenarios[26] to see how "Shared Infrastructure Services" elements might contribute to them, and consider how to make that contribution more explicit. Andy Powell's 2005 papers, *The JISC Resource Discovery Landscape*[27] and *A 'service oriented' view of the JISC Information Environment*[28], were also useful in this respect.

These scenarios could be presented in other ways, depending on the intended audience and purpose in any particular case. Comparative scenarios could be provided, showing how much more labour-intensive it is to carry out these activities or achieve these goals without the underlying support of shared infrastructure services. For a more technically-orientated audience, the scenarios could be presented in a two-column layout, with the underlying machine-to-machine activity presented against the human activity steps. This would require a much more detailed breakdown of the human activity, as the underlying complexity that would be described is hidden from the human actors so that their tasks are simplified as much as possible. The work involved is beyond the scope of the current report, but could be worth undertaking in the future.

In most cases, the shared infrastructure services are not directly seen by the actors, but are running behind the scenes, linking together or being called upon by the tools and services that the actors use to perform their roles. As in real life, actors – the imagined people in scenarios – may have more than one role, although each scenario typically focuses on a single role. In most of the scenarios presented here, the imagined people and their roles are:

- **Sudeep** is a member of staff of the University of West Essex Geography Department. A post-doctoral research scientist investigating the human geography of alluvial flood plains, Sudeep is also a lecturer with responsibility for the teaching and supervision of postgraduate degree candidates, and is a course leader for an undergraduate Area Studies programme.
- **Lisa** is a first year undergraduate student at the University of West Essex taking a course module in the Area Studies programme.
- **Raymond** is a member of staff of the University of West Essex library. Raymond is Assistant Librarian with responsibility for digital resources and collection development.
- **Helen** is an RAE Support Officer at the University of West Essex.

7.1 Deposit in repositories and a Virtual Learning Environment (VLE)

Sudeep has created an e-Learning package for the review and self-assessment of basic concepts in Area Studies. In order to deposit it in his institutional repository for learning objects and in JORUM, and make it available to his students through his institution's VLE, Sudeep submits it in a single process using a utility provided in his institution's portal. The deposit utility includes a simple form that collects basic information about the format, provenance, intended audience and IPR status of the deposited item, and a choice of locations for deposit. Sudeep selects a date range for which the package is to be made available through the VLE, beginning with the current date, and selects the course for which it is to be used, so the package is immediately available to the students registered for that course and a notification email is sent to them. (This scenario could be enriched by showing how creation of the learning package was supported by shared infrastructure services.)

Raymond processes the latest additions to the University of West Essex learning object repository, which includes the e-Learning package deposited by Sudeep. Using the information extracted from the item itself and from Sudeep's form, Raymond edits the metadata record for the item. To enrich the subject terms, he confirms the accuracy of the automatically generated mappings from the specialist geographic scheme used in the Geography Department to the more general scheme used for the institutional repository. He also checks that preservation metadata has been generated, and then accepts the output of records in the two different application profiles required for Jorum and for the university's own institutional repository.

These activities are possible because behind the scenes, facilitated by shared infrastructure services:

- Depositing of material is supported in the institution's own portal by a metadata schema registry through sharing of information about metadata schemas and application profiles, and also by application profile creation utilities (see **Related metadata schema registry scenarios**, below);
- IPR information sharing is supported by an extended Romeo service, and perhaps by a license registry, which might be a further extension of the Romeo service, or a service based on ONIX for Licensing Terms, or a combination of the two;
- Semi-automated metadata creation is supported by terminology and other services, including for example:
    - Creator/author name indexing is supported by a name authority service;
    - Subject indexing enrichment is supported by a combination of a geospatial (gazetteer) terminology service and a more general terminology mapping service;
    - Preservation metadata creation is supported through services layered on a representation information registry, format registry and persistent identifier service.

Related metadata schema registry scenarios. For example, a schema designer for a learning object repository makes the metadata application profile created for the repository available – to content providers, presentation services, and the developers of other learning object repositories – by submitting a description of it to a metadata schema registry. For a detailed description of this scenario and others related to the design and use of metadata schemas, see: *Usage Scenarios for the IE Metadata Schema Registry*[29].

7.2 Create/maintain a reading list; use a reading list

Sudeep edits an Area Studies reading list, using a reading list utility provided within the university's portal. The reading list includes journal articles, books and book chapters, and web pages – a mix of digital and physical resources. A former student on the Area Studies course has sent him copies of an article and a paper she has written. As they are both relevant, Sudeep adds references for these to the reading list, although he knows that the journal publishing the paper is one to which the university library does not subscribe. He also adds the former student's thesis to the reading list. Sudeep adds another item, a new edition of a book that is already on the list. An email alert informing the library that the reading list has been updated is automatically generated. Sudeep adds a note about requesting that the library subscribe to the journal, and confirms the sending of the email alert.

Raymond receives the update alert, and uses a facility linking the VLE with the catalogue database to create a link with the library's catalogue record for the new edition of the book, which is now available in the library, and with the circulation module, which enables students to request the book on loan direct from the reading list. He changes the loan status of the book from normal to short loan. Raymond also adds a link to the thesis in the library's electronic theses collection. He adds Sudeep's request for a journal subscription to a list to be considered at the next collection development meeting. An automatically generated email alert informing them that the reading list has been updated is sent to students on the Area Studies course, and also to any students or staff of the university who have added to their profiles an interest in Area Studies.

Lisa is a student on an Area Studies module taught by Sudeep. She remembers when she receives the alert that she wants to consult some of the items on this reading list for her next assigned essay. Lisa logs on to the library portal through a link in the alert message. She decides to look at the new items first, and when she requests the new edition of the book, is informed that she can collect it from the library six days from today. The article, published in an open access journal, although interesting is short and refers to the full paper for further detail. When Lisa realises that she will need to read the full paper, she accesses it from the link in the reading list. The paper is available from the author's institutional repository as well as from the online journal publishing it. The publisher's license terms determine that the paper is under embargo outside the author's own institution for a further seventeen days, a week after the deadline for her essay, and the University of West Essex does not subscribe to the journal. However, Lisa has a student membership in an association that entitles her to access to the journal, so she is allowed to access the paper direct from the journal's web site.

These activities are possible because behind the scenes, facilitated by shared infrastructure services:

- Authentication and authorisation services support a single sign-on procedure for access to all the resources;
- Personalisation services support authentication of individual's access rights based on characteristics in addition to affiliation to the educational institution;
- Rights management is supported by an extended Romeo service, and perhaps by a license registry, which might be a further extension of the Romeo service, or a service based on ONIX for Licensing Terms, or a combination of the two;
- Persistent identifiers are supported for items on the reading list;
- OpenURL resolver services support directing of user to an appropriate copy.

7.3 Search for scientific data; add data to repository

Sudeep uses a specialist geospatial data search facility provided through his local portal to search for any recently published geographic data related to research he is currently conducting. He finds one dataset that provides additional statistical detail related to changes in population size. This can be used in conjunction with data he has been collecting from water-level measuring instruments and historic records. The description of the dataset includes links to two published articles based on research that used the dataset, and to a research report in an institutional repository. Sudeep retrieves these for reference, as well as the dataset. He deposits his new/derived/value-added geographic data set. Representation information is generated semi-automatically to support data preservation/curation.

These activities are possible because behind the scenes, facilitated by shared infrastructure services:

- A service registry has previously been used by the portal to locate suitable collections for subsequent searching;
- Terminology services (for geospatial terminology mapping and non-geospatial keyword mapping) are used by the geospatial data search facility to support more comprehensive result set retrieval, by enabling search across dataset metadata for equivalent, near-equivalent, or related terms;
- A representation information registry supports semi-automated creation of preservation metadata.
- Persistent identifiers and OpenURL resolvers support trust based on assurance of authenticity and appropriate retrieval routes;
- A metadata schema registry has supported the creation of appropriate application profiles for description of resources

7.3.1 Related search and deposit scenario for humanities data

A dance student locates recorded music, music scores, choreographic notation, and still and moving images. (Sources for material have been pre-identified through the portal calling on a services registry.) The student uses the notation to link the choreography he creates to the music, and deposits the result in the VLE for other dance students to access and perform. The performance is videoed, and the student adds the video to the previous deposit, tagging it as ready for assessment. Note that this scenario could also be used to illustrate searching across types of resources.

7.3.2 Related search, data creation, manipulation and deposit scenario for scientific data

Provided by Stephen Rankin of CCLRC, building on the scenario outlined in the JISC IE Architecture Usage Scenarios referenced above.

A post-doctoral researcher in biomedicine is interested in establishing links between environmental air quality and a specific lung disease. The researcher begins by performing a cross-search of three selected sites, using the same keywords that are mapped automatically to a common biomedical thesaurus, via the BIOME subject gateway. The search covers the traditionally published primary literature directly via Medline, the BioMed Pre-print archive and the unpublished databank of epidemiological data held by the Medical Research Council (MRC) that is located on the "Grid". Access to each of these resources is managed by the ATHENS/SPARTA authentication mechanism. The researcher retrieves a number of bibliographic references to journal articles and then immediately downloads the full text of the ten most recent ones. A number of pre-prints by the same authors are retrieved and three sets of raw data from the MRC data repository are collected.

The researcher then selects relevant numerical results data from each of these sources and retrieves the representation information for the raw data from a representation information registry/repository located at the MRC. The representation information allows the researcher to obtain the relevant structure information about the raw data file format (possibly an EAST

description) and also semantic information in the form of a data dictionary (possible a DEDSL data dictionary) that describes their content (units, value descriptions etc). They also obtain a statistical software package that is EAST/DEDSL aware, so immediately they can start to understand their content and run tests on any valid associations between the data sets.

Digging deeper into the representation information network, the researcher discovers information describing EAST and DEDSL (in a representation information registry/repository located at the CCLRC) and their associated software tools/libraries. They also discover more documentation describing the typical domain-specific processing techniques for the raw data (algorithms) that gives them the ability to start to write a specific data processing application relating to their own research. Given that the shared infrastructure services provide client application level access, the researcher can enable their application to automatically pull down new raw data and process it producing new results. Finally, the researcher saves the search to their user profile for repeating at regular monthly intervals.

7.4 Search across subject areas, types of resources, and sectors beyond higher education

Lisa needs to go beyond the reading list to further her learning. Area studies is a cross-disciplinary field. When Lisa enters terms she thinks of in the library portal search facility, results are returned to her that exactly match her terms, but also results that match closely-related terms across a number of indexing schemes from a variety of general and specialist indexing schemes. She is also offered an option to refine her search through disambiguation of one of the terms. The refined results list provides her with forty-seven possibly useful resources. These include journal articles, books, web sites, digital images, objects in museum collections, items in public library local history collections, and online course materials.

Lisa checks some of the resources whose descriptions look most promising. She finds an interesting, though brief, reference to her topic in online course material from the Open University. By following links from the citation, she finds other relevant material by the author cited, and retrieves a research report from a social science subject-based repository.

Lisa must consult additional resources relevant to circumstances in Hartfield, a nearby town she is using as an example in her essay. Lisa wants to consult Hartfield town plans and strategic planning reports for the local authority within which it lies. She discovers that the Hartfield Public Library has a collection of town plans, and that the county Record Office holds archive copies of council reports. Lisa notes the location, contact details, and hours of public access to these collections, in order to follow up by making arrangements to visit the collections.

These activities are possible because behind the scenes, facilitated by shared infrastructure services:

- Terminologies services support disambiguation and cross-disciplinary searching by mapping terms across various schemes;
- Name authority services support the identification of authors and organisations across institutional and other types of repositories.
- A service registry with collection descriptions based on metadata standards work of the Collection Description Focus has supported the previous identification via the library portal of appropriate collections.
- The collection descriptions available from the service registry are available (perhaps on-the-fly) for display to searchers.

7.5 Access to institutional repositories for administrative purposes; other administrative and marketing benefits

Helen is preparing the universities RAE submission. Some academic authors have identical or similar names to other authors, and many authors use different versions of their name at different points, so disambiguation of identical or similar names and the creation of name authority records may be required. The appropriate version of each publication for submission

must be identified. Helen needs to have specific formats for the submission, and these are generated automatically for cases where a researcher/author has deposited a publication in a different format.

These activities are facilitated behind the scenes by shared infrastructure services including:

- a format conversion service
- services supporting version management/control
- identifier services
- a name authority service or services

7.6 Other potential scenarios

The scenarios set out in this report have focussed primarily on discovery and curation activities that could be facilitated by shared infrastructure services. However, it is recognised that shared infrastructure services have potential use across all the areas covered by the e-Framework, including the support of administrative processes. Additional attention should be given to the potential benefit of shared infrastructure to content providers (both commercial and non-commercial) that make their resources available for use by the JISC community.

8 Personalisation

Personalisation is defined as: 'the ability of a Network Service to be shaped or re-shaped so as to better meet the individual needs or wants of a user' (adapted from O'Looney[30]).

'Personalization involves a process of gathering user information during interaction with the user, which is then used to deliver appropriate content and services, tailor-made to the user's needs. The aim is to improve the user's experience of a service.'[31]

It is notable that definitions focus on improving services to users. There are many different ways of achieving this, with a range of approaches potentially falling under the banner of personalisation.

A report on *Personalisation in presentation services*[32] was commissioned by JISC in 2004. It distinguished between Customisation, where the users have responsibility for customising their own experience, and Adaptive Personalisation where the availability of options, interface, access or functionality is based upon knowledge about users gained from tracking user activity and/or other sources of user information. It was found that most JISC services seem to be providing customisation rather than adaptive personalisation.

Adaptive Personalisation is usually based either on collaborative filtering (allows a service to identify items of potential interest to a particular user based on the preferences of other users with similar characteristics and/or activity records, e.g. Amazon) or rules base filtering (based on preset rules about relationships between items and user profiles, e.g. A is a PhD student, therefore has access to resources b and c)

From a shared infrastructure point of view, services that personalisation functions are likely to need to access include the Institutional Profiling Service (e.g. to determine access rights as a member of an institution), Authentication/authorisation Service (e.g. to determine an individual's access right to a particular resource), Service Registry (e.g. to enable a user's portal to determine what collections are available to end-users, which could then be used to personalise the user information environment). Personalisation may occur either at an individual level, or group level, so access to administrative data may also be required, e.g. to access course or module data so that resources can be targeted by subject and level.

[Note: The Access Management Federation is based on the trust relationship between Identity Providers (IdP) and Service Providers (SP), using Shibboleth technology. Responsibility for user authentication is devolved to the user's home institution.]

The information environment experienced could be determined firstly by authentication/authorisation and then by user customisation – e.g. choosing preferred resources

within authorised boundaries**.** Users may also wish to customise the user interface, choose font size and preferred delivery formats, e.g. for accessibility purposes.

VLEs and JORUM-like repositories will be useful in providing:

- Accessible versions of resources (e.g. electronic braille file of a text item)
- Alternative resources (where the original cannot be used even when transcribed to another form, an alternative learning object or resource that fulfils the same learning objectives could be provided)

Within the context of the IE architecture diagram, personalisation can be seen as a function invoked by components at the presentation layer, in particular portals. However personalisation of 'push' services can also apply at the fusion and provision layer. Personalisation may be embodied entirely within the portal or may rely on interaction with shared infrastructure.

The JISC report concludes that personalisation can improve efficiency, reveal inadequacies in business processes and allow services and learning materials to be effectively targeted. Although personalisation is no substitute for user requirements analysis and user-centred design, accessibility to users of all abilities may be improved by offering options such as switching off graphics, or changing font-sizes or background colours – all Web sites should consider this. True personalisation is more than allowing users to 're-skin' the interface or change the position of screen elements.

The report does not recommend setting up national services for personalisation or user profiles and it discourages the development of national standards in an area where international de facto standards are still developing.

JISC are currently taking this work forward in a number of ways. A Call was put out within the e-Infrastructure Programme in April 2006 for identity management work. The JISC IE Programme Team has also held a recent cross domain personalisation workshop to investigate some further actions on how best to address personalisation. This work has not yet reported but will by October 2006.

The option of holding learner profiles centrally needs to be considered, since the aim is to pass profiles from institution to institution as the learner moves through the educational system (and can be used by the learner as the basis of CVs and applications for further courses and jobs). Learner profiles implementations are read/write by both student and staff (the interfaces restrict which areas each user type sees and the interfaces can be personalised in terms of screen font, colours, etc). They record elements such as deposit of work for assessment, record of tutorials, assignment completion and assessment, exam marks/grades etc.

- ***Recommendation: JISC should consider looking at whether shared infrastructure services could/should provide additional support for personalisation in order to enable more adaptive personalisation within the IE.***

9 Repositories

Repositories hold content of various types and so form part of the provision layer of the JISC Information Environment. A number of repository projects have been funded under the JISC Digital Repositories Programme. The current Programme has been informed by two key documents: the *Digital repositories review* by Rachel Heery and Sheila Anderson in 2005 [33] accompanied the programme call, while the 2006 paper *Digital repositories roadmap: looking forward* [34] by Rachel Heery and Andy Powell contrasts the current situation with the authors' vision for 2010.

It is not in the remit of the current study to look in detail at individual repository projects; however, as noted in the above-mentioned reports, repositories will need the support of the shared services infrastructure in order to deliver to their full potential.

Two of the characteristics of a digital repository are that repository architecture manages content as well as metadata, and that it offers a minimum set of basic services e.g. put, get, search, access control. These are areas where the support of the infrastructure will be needed.

The *Digital repositories review* notes that:

> *"The intersection of interest across domains offers possibilities for various crossovers of technologies. There is also potential for sharing experience, sharing tools and undertaking collaborative development work. It is important that there is coordination of this activity and that an appropriate level of interoperability is achieved, without placing barriers on innovative work."*

The review comments that development and deployment of repositories is currently patchy and immature and that future services will need to rely on well-structured work-flow between repositories, and on interfaces between repositories and other components of the information environment. It also makes the point that for the user, the priority is to gain access to the information they need, while the corresponding challenge for the programme is to build repository content and deliver the benefits of repositories without burdening content creators and end-users with any additional process.

The *Digital repositories roadmap* report envisions the situation in 2010 as follows:

- Repositories will be much more interoperable with systems used to support learning and teaching, Virtual/Managed/Personal Learning Environments, assessment systems, ePortfolios, etc., as well as with authoring tools, other repositories, portals and library systems.
- Repositories will support aggregation of content (both metadata and full data) by service providers.
- Repositories will consume services such as content (or metadata) enrichment services.
- Users will be able to discover, locate, access and use geospatial content that is distributed across institutions and organisations more seamlessly, ideally through an integrated and interoperable services layer.

In order to make this vision a reality, a number of challenges must be overcome. Some challenges are cultural and outside the remit of this study – mandating that research outputs are made available in open access repositories, academics automatically depositing papers and/or learning objects and repositories being embedded within institutional strategies. Other challenges are technical, and JISC shared infrastructure services may form part of the solutions.

JISC recently funded a scoping study on Linking Repositories and a report from the study is now available[35]. A presentation on the Linking Repositories scoping study was given by Alma Swan at the Integrating Infrastructure cluster session at the 2nd JISC Digital Repositories Programme Meeting, 27-28th March 2006, Warwick[36]. Discussion raised the issue of what other services need to be layered onto repositories; it was felt that some services would be needed at an institutional level, while others (e.g. preservation) would be layered above that, perhaps at the collaborative/aggregated level, which could offer economies of scale. Participants put together the following list of potential services:

- RAE submission
- Personal CV
- Group/departmental/faculty biographies
- Overlay journals
- Preservation services
- Access to multiple file formats
- Accessibility
- Linking data and publication

Note that simply specifying that, say, preservation services are required is not sufficient. Preservation services might require a range of infrastructure services – representation

information, format registry, name authority, geospatial services, metadata schema registry – and this needs to be defined in more detail.

Some Heery & Powell recommendations on repositories would be equally applicable in the context of shared infrastructure services, as follows:

- Agree the machine-to-machine interfaces (the services) that open access repositories should support in order to ingest and make available content and metadata.
- Work towards DRM solutions that allow software to take decisions based on machine-readable licences.
- Develop modular services to be provided by repository software suppliers which can be plugged in to deliver different functions e.g. preservation, RAE outputs, personal profiles (CVs), etc.
- Additionally there would be benefit in a repository junction (e.g. OpenDOAR, IESR) which would enable would-be depositors to identify an appropriate place to deposit (e.g. an academic whose institution doesn't yet have its own repository – use a subject one or an interim one). This might be designed at human interface level first, and then M2M (cf. Romeo).

10 E-Science Grid

The way that research is carried out is changing, with a growing emphasis on large-scale distributed global collaborations that are enabled by the Internet. Typically, such collaborative enterprises require access to very large data collections, very large-scale computing resources and high performance visualisation back to the individual user researchers.

The World Wide Web has provided access to information on web pages in html, but a much more powerful infrastructure is needed to support e-research. In addition to information stored in web pages, researchers will need easy access to expensive remote facilities, to computing resources and to information stored in dedicated databases.

The e-Science Grid[37] is an architecture that has the potential to bring all these issues together and make a reality of the vision. Increasingly the Grid is viewed as a 'web service with extras on top' and will use shared infrastructure services to help deliver its full potential, so development in the e-Science area will take account of what is developed by shared infrastructure services. Just as shared services will benefit the grid, the increased compute services of the Grid will benefit the JISC community.

The Grid will potentially make use of all shared infrastructure services currently in development. Given its focus on large-scale enterprise, it is important that services that are used by the Grid are robust and scalability is an issue – will service X function with a million hits a day? Although not specifically noted in the JISC IE architecture diagram at Figure 2, the desktop browser or user could also be a plug-in application, so services will need to interoperate with such application interfaces. Working to national and, where available international, standards is also a key issue.

The interest in and relevance of Grid technology is not restricted to the UK and there are some initiatives in the international arena. It will be important that the UK works with any evolving / developed European infrastructure and is aware of what comes out of events such as the 2nd Concertation Workshop on eInfrastructure held in 2005[38] and the Task Force on the Permanent Access to the Records of Science (n.b. in the European context 'science' includes humanities research)[39]. The European Strategy Forum on Research Infrastructure (ESFRI)[40] is preparing a European Roadmap for new research infrastructures of pan-European interest, while in the USA, EPSCoR[41] is developing a roadmap for cyber infrastructure for large-scale science and engineering. The Internet Engineering Task Force[42] is looking at obtaining general consensus that can be developed into guidelines for interoperable implementation. The National Science Foundation (NSF) Cyberinfrastructure report[43] is another useful resource.

11 Current Status of the Shared Infrastructure

This section reviews the current status of JISC shared infrastructure projects, as well as some UK and international initiatives that complement JISC work or fill gaps. The sub-sections are principally based on the JISC shared infrastructure as illustrated in section 5 above (the JISC Information Environment).

The JISC Information Environment and the Shared Infrastructure Services also form a part of the supporting structure for the new JISC e-Infrastructure Programme[44], which formally commences in September 2006. e-Infrastructure embraces networks, grids, data centres and collaborative environments, and can include supporting operations centres, service registries, single-sign on, certificate authorities, training and help-desk services. The integration of these defines e-Infrastructure.

Given the importance of standards across the shared infrastructure, it is also worth highlighting the JISC Standards Framework[45] which is developing a layered approach to the selection and use of open standards in order to support development work within the UK higher and further educational communities. A standards 'catalogue' is included. It acknowledges that there is not a universal solution, but rather the need to recognise local, regional and cultural factors which will inform the selection and use of open standards. To place the layered approach in context, case studies are provided of the types of environments in which the standards framework can be implemented.

11.1 Identifier Services and open linking

An Identifier Service is a network service that maintains and provides an association between an identifier and some metadata about the identified resource. Typically, an identifier service takes an identifier of a resource and returns a locator for it (usually in the form of a URL).

11.1.1 Persistent identifier services

A persistent identifier service is one that is specifically designed to offer consistent services over very long periods of time. It will enable long-term access and re-use of content by end users and other systems. However there is a wide range of different schemes available. The best known include ARK, DOI, Handle, ISBN, ISSN, PURL, URI, URL, URN, but there are many more. There seems to be little consensus about why one system should be chosen over another, and what benefits and pitfalls each approach brings.[46]

The long-term adoption of identifier systems is dependent on a complex mix of political, social, financial and technical issues. Identifiers cannot hope to achieve persistence unless they are widely adopted within digital library services.

In addition to identifying data, there is likely to be wider application:

> *Persistent Unique Identifier (or an alternative means to achieve this functionality) will enable global cross-referencing between data objects. Such Identifiers will not only be used for data and software but also for other resources such as people, organisations, etc.*
>
> *On the other hand, any scheme of identification is likely to undergo evolution so preservation, and in particular integration of archival and current data, is likely to require active management of identifiers.* [47]

11.1.2 OpenURL

The OpenURL standard is a syntax to create web-transportable packages of metadata and/or identifiers about an information object. Such packages are at the core of context-sensitive or open link technology. The OpenURL is needed because conventional web links do not take into account the identity of the user: they take all users to the same target. This causes some

problems. For example, when more than one institution provides access to copies of the same electronic article, the link from citation to full text should resolve to a copy that is accessible to the user[48].

The original version of OpenURL, now designated OpenURL version 0.1, provided both a common linking syntax and a solution to the appropriate copy problem[49]. The OpenURL concept was developed as part of a research project (called SFX 'special effects') by Herbert Van de Sompel and Patrick Hochstenbach at Ghent University. It was then acquired by Ex Libris which currently sells the SFX OpenURL resolver.

- Many major discovery resources now provide OpenURL source links, as do some journal publishers. But there are still some major resources that do not provide OpenURL source links or enable target linking into their content. This situation causes concern to librarians who view such resources as significant gaps in the 'joined up' linking experience that they would like to provide to their readers using the OpenURL technology in which they have invested[50].

11.1.3 OpenURL resolvers

SFX[51] from Ex Libris is the UK market leader in commercial OpenURL resolver products. While there are several other products in use in the UK (including the resolver developed by the ZBLSA project at EDINA), SFX is mentioned here because of its wide use. SFX operates independently from integrated library systems, so libraries do not have to be MetaLib users.

SFX allows context-sensitive linking between Web resources. It uses the OpenURL standard for interoperability between information resources and service components. For users whose institutions have access to the SFX link server, a library-defined SFX button appears with each retrieved reference, whether the resource is hosted locally by the institution or remotely by a third party.

OpenURL resolvers are very expensive to purchase and implement. This is creating a situation where there is a difference in user experience at the 'have' and 'have not' institutions. Services such as Zetoc and OpenURL router do try to provide some default resolution on links to other search engines but it is not the same as appropriate linking. The current default LinkFinderPlus service provided to UK HE/FE via Zetoc and the router, will be withdrawn in November 2006 for funding reasons. It would therefore be useful for JISC to provide some form of default OpenURL resolver for smaller/poorer institutions; OCLC Openly Informatics[52] have expressed interest in being involved.

In addition to start-up costs, there is also a lot of ongoing work for librarians to maintain their resolvers - they cannot be set up and expected to run without maintenance.

11.1.4 OpenURL Router

The OpenURL Router[53] is an operational service linking different bibliographic services, typically an abstracting and indexing database (a referrer) and services (resolvers) which locate copies relating to the reference that a user has found in his or her search. The OpenURL Router is an offshoot of the ZBLSA[54] project, based at EDINA. It is provided to all HE and FE institutions in the UK.

The OpenURL Router works by offering a central registry of institutions' OpenURL resolvers. An institution registers details of its resolver just once. When the resolver has been registered, any service provider can provide users from that institution with OpenURL links to their resolver.

Services such as Copac, which are free to use, cannot determine a user's institution, and hence their resolver. The facility to send OpenURLs to the Router is therefore a significant benefit.

The coverage of the OpenURL Router is UK HE/FE only. This may cause problems for services with a wider or non-matching coverage.

11.1.4.1 What does the project aim to deliver?

- Help to institutions with OpenURL Resolvers to establish OpenURL links from a wider range of services

- Enable providers of OpenURL aware services to link to the appropriate OpenURL Resolver for each of their end users.
- Extend the range of services in which OpenURL links can usefully be deployed.

11.1.4.2 Current status

In full service

11.1.4.3 Support function within JISC IE

The OpenURL Router enables providers of OpenURL-aware services to link to the appropriate OpenURL Resolver for each of their end users.

11.1.4.4 Risk assessment

Without the OpenURL Router, linking requires configuration of links between each institution with a resolver and each service provider on a pairwise basis. Each institution must arrange for links to be set up with each service to which they subscribe, and if the resolver is changed, this process must be repeated. Each service provider has to maintain tables mapping each of their end users to their institutions, and institutions to resolvers.

Another risk is that institutions will not bother to register their resolvers with the router. Its coverage is currently poor, which makes services reluctant to use it.

The router does not have an OAI-PMH interface (although there are alternative methods of getting the data currently). This would be useful to allow OpenURL source services to harvest the data to part-populate their own internal registries.

11.1.5 OpenURL Gateway

OCLC have developed the OpenURL Gateway[55], which is very similar in concept to the OpenURL Router, although it appears to be based on a richer set of information about the capabilities of the OpenURL routers that it knows about. The resolver's other difference is that its knowledge is worldwide rather than UK only. Like the OpenURL Router, the OpenURL Gateway sits between OpenURL sources and OpenURL resolvers, in order that each source does not have to maintain knowledge about each end-user's preferred OpenURL resolver.

In the UK, the knowledge about available OpenURL resolvers is maintained by the OpenURL Router itself (although there is also potential for IESR to include OpenURL resolvers). It would be sensible for this knowledge to be shared with OCLC's OpenURL Gateway, so that if a UK user inadvertently ends up at the OCLC OpenURL Gateway rather than the EDINA OpenURL Router, it can still do something sensible for them. However a global solution should be found to the problem of seamlessly discovering the correct OpenURL resolver[56].

The CoinS specification[57] shows the beginnings of approaches that allow a user to control/choose which resolver to use.

11.1.6 JISC Future Strategy

- ***Recommendation: an OAI-PMH interface for the OpenURL Router would be useful, allowing OpenURL source services to harvest the data to part-populate their own internal registries.***
- ***Recommendation: JISC/OCLC/Digital Library Federation (DLF) should cooperate on a global solution to the problem of seamlessly discovering the correct OpenURL resolver.***
- ***Recommendation: JISC should consider the provision of a default OpenURL resolver for smaller/poorer institutions, in collaboration with a partner such as OCLC Openly Informatics or EDINA (possibly based on ZBLSA work).***

11.2 Representation Information and Format Registries

A quotation from a briefing paper on *File Format and XML Schema Registries* by Alex Ball[58] succinctly sets the context for this area of the JISC Information Environment.

Digital files are, fundamentally, strings of binary digits (bits). In order to process them, one must know the format they are in, and further, what software is needed to read that format. Even after the file has been successfully opened, extra information may be needed in order to fully understand the contents. In the terms of the Open Archival Information System (OAIS) Reference Model, the information required to transform a stream of bits into something intelligible is called 'representation information'.

Providing representation information so that digital files remain perpetually intelligible is a significant challenge for the JISC Information Environment. The OAIS Reference Model[59] mentions using representation networks in order to reduce the burden on individual archival information systems; i.e. holding the representation information in an external source known as a file format registry. Repositories will then be able to reference the database whenever a file in a new format is added, and any new information on a file format needs only to be added once. (An example of this would be any new version of the Word document file format, which already exists in a number of versions.)

Representation information includes more information than simply file format data; it also includes semantic representation information, such as instrument calibrations, data units and other information necessary to interpret scientific data. For example, it would not be sufficient to know that something was an SPSS (statistics package) file; it would also be important to know the variables that had been used.

File format registries are databases of representation information solely concerned with file formats. They can either be maintained as a node in a wider network (this form is easier and cheaper to maintain but may encounter difficulties if the specification for a format becomes unavailable) or they can attempt to operate as a terminus, collecting and preserving copies of specifications and software in a local repository.

A number of privately-maintained websites contain information about file formats although these are aimed more at the casual user and are likely not to be maintained in the longer term. Examples are Wotsit's Format[60], the File Format Encyclopedia[61], and FILExt, 'the File Extension Source'[62].

Initiated in 2000 in the US, the National Digital Infrastructure and Preservation Program (NDIIPP)[63] aims to develop a national strategy to collect, archive and preserve the burgeoning amounts of digital content, especially materials that are created only in digital formats, for current and future generations. The Program will seek to provide a national focus on important policy, standards and technical components necessary to preserve digital content. Investments in modelling and testing various options and technical solutions will take place over several years, resulting in recommendations to the U.S. Congress about the most viable and sustainable options for long-term preservation. It is noted that similar work has begun in Europe with the formation of EDIIP (no reference located).

One tool that has been developed within NDIIPP is a digital formats registry. The Library of Congress created the web-based resource *Sustainability of Digital Formats: Planning for Library of Congress Collections*; this is primarily concerned with providing advice on the suitability of particular file formats for long-term preservation[64]. The records held within this resource are designed for human-readable access, and are therefore unsuitable as a basis for automated tools.

11.2.1 DCC Representation Information Registry / Repository

The Digital Curation Centre (DCC) has developed a Representation Information Registry / Repository (RI RegRep)[65]. The DCC sees Representation Information as a key concept for information preservation and a (distributed) Registry/Repository of Representation Information is an essential service. This is geared to the needs of the e-Science community, and in particular for curating experimental data sets. As a registry / repository, it holds details of file formats and copies of format specifications and rendering software. It will also hold semantic representation information, such as instrument calibrations, data units and other information necessary to interpret scientific data.

Both registry and repository elements were needed and it was decided to use the existing 'industry standard' ebXML, for which a number of tools have been developed, rather than re-

invent the wheel. The implementation used 'free ebXML' but is not restricted to continuing to use it (an alternative could be used in the future); layers were then built on top using JAXR.

Identifiers are required for the repository to locate and return something. However, it is felt that identifiers are inherently unreliable and the RI RegRep will not be creating any identifiers of its own. It will input as many identifiers as are known for something in the hope that in the long term, at least one will still function.

Registries need to allow searching so someone can ascertain whether someone else has already described the data, in such a way that allows re-use of that description. Versioning will be important – it is not enough to simply know the name of a format as these often go through cycles of development with several releases over time.

DCC work in this area is now closely linked with the EU project CASPAR (Cultural, Artistic and Scientific Knowledge for Preservation Access and Retrieval)[66], which aims to develop testbeds in such a way that they can be embedded in partner data services.

11.2.1.1 What the project aims to deliver

The *DDC Approach to Digital Curation*[67] includes the following aims and objectives:

- The Registry/Repository must itself be an exemplar trustworthy OAIS repository, for long-term preservation of the Representation Information which it holds, and it will be OAIS certified in due course.
- A number of tools and services can be built on the Registry service itself and on the Repository's design and implementation.
- The tools and services will aim to promote interoperability and automated use as far as possible, and will support information preservation over the long term as well as current and future usability of information held in repositories of all kinds.
- The tools and services must be easily integrated into many of the other UK and international projects which are addressing the issue of digital curation.

11.2.1.2 Current status

Project with identified service potential

The RI RegRep is already in use; it has been populated with some basic data, which is being 'tidied up'. By the end of June 2006, it should be ready to take a lot more data in preparation for participation in CASPAR. There is a need to develop more tools for data entry and editing.

At present it is best to have local caches of the RI RegRep to be used with local applications; at a later stage it should be possible to rely on the e-infrastructure. For the future, it must be able to support scaleability and have sufficiently robust production strength for CASPAR work.

The design of the registry as an implementation with layers on top that support further applications means that there is a fairly thin API that should be enable M2M interoperability with the PRONOM registry. Over the next 3 to 4 months it will be tested with several applications that can talk to it using the API.

File conversion services could be built on top and DCC is looking at using the EAST description language in developing such services.

The aim is to work with an inner circle initially and moving out to other data holdings once there are implementations and applications that can be demonstrated. Initial work will focus on the easier areas (social sciences) and then move to the more complicated ones (e.g. genetics).

11.2.1.3 Support function within JISC IE

RI RegRep and its services provide two means of support. Firstly, it is intended to support long-term preservation; at the moment the user might just need to know that data is in Word format but in 50 years the user might also need to use a Word emulator to access the data. However, it is important that funding for new work is not compromised by the cost of preserving existing data.

Secondly, it could also support new ways of accessing data via virtualisations for the Grids. For example, archaeological artefacts may be scanned in various ways – neutron scans, laser

scans and xray scans. If the data format from each of these scanning methods is recorded using a description language, then it would be possible, say, for the data from an xray scan to be used to generate the equivalent of a laser scan.

11.2.1.4 Risk assessment

Without this type of registry, significant amounts of information from preservation repositories would be lost; initially there would be little impact but this would increase over time with some forms of digital resources becoming inaccessible. The risk to the JISC IE would be that while alternative strategies such as using say, GDFR, TOM and PRONOM in combination, might enable recovery of much data, such methods would require significant effort and funding. It is noted that JISC is currently funding little in the way of preservation repositories.

There is also the risk to grid development work if the JISC IE is not able to offer the format conversion services that would enable virtualisations.

There is some other work and interest in this area both nationally and internationally. The RLG Audit makes passing references to registries, while the GDFR appears to intend spending a lot of time writing a registry, and seems not to have tackled the identifier issues apart from having some sort of classification. In Italy some new laws on document preservation will have implications for architects – and they will need to start working in this area. In the UK, PRONOM is operational but at present only has flat web pages, thus limiting interoperability.

At present there is no alternative to the RI RegRep available. The JISC IE would be limited in the support it could give to the Grids.

11.2.2 PRONOM

PRONOM is a format registry being developed and maintained by The National Archives (TNA)'.[68] (the TNA was previously the Public Record Office (PRO). It is an online registry of technical information about file formats, software and digital preservation related tools; in addition to the current human-readable interface, there are plans to develop M2M capability to allow automated data exchange with other databases and to interoperate with various automated services. Web services interfaces, probably complemented by REST and OAI-PMH interfaces, should be available within the next 12 months (i.e. July 2007).

It was developed as service for the TNA Digital Preservation department, which has been operating a digital archive for born-digital public records since 2003. The database currently holds detailed information on over 550 current and obsolete file formats, together with the software that is required to access them. Developments are planned for PRONOM under the Technology Watch project in the TNA Seamless Flow programme.

PRONOM holds information about various classes of representation information, at the sub-format level (as with DCC), such as compression algorithms and character encoding schemes, and at higher levels, such as software tools, operating systems and hardware platforms. The differences between DCC and PRONOM are mainly due to the different focuses, i.e. scientific data sets need to be described in different ways to office and other formats.

One of the support functions identified for development is the ability to generate migration pathways between formats – a format conversion service – which would enable a delivery service to transform an electronic document requested by a user into a format they could use.

TNA has developed the Digital Record Object Identifier (DROID)[69], a Java tool for automatically identifying file formats, using a signature stored in PRONOM.

The PRONOM Unique Identifier (PUID) provides persistent unique identifiers for file formats recorded in the registry, and has been adopted as the preferred encoding scheme for file formats with the e-Government Metadata Standard. The DCC RepInf Reg/Rep intends to re-use the PUIDs where it can under its multi-identifier strategy.

11.2.2.1 What the project aims to deliver

The primary aim is to provide a service for TNA digital services work, supporting both passive digital preservation (storage) and active preservation (maintaining access over time). However,

TNA also recognises the value of such a service outside its own activities, and is actively working to make it more widely usable.

Stage 1 will deliver a new release of PRONOM; expected release date is c. April 2007.

Stage 2 will further develop PRONOM so that other services (web services interfaces and M2M interfaces) can be built on the registry. Expected completion date is c. April 2008.

TNA is contributing to the JISC Preserve project, looking at how repositories can make use of third party preservation services. The project is acting as a testbed to investigate the embedding of the DROID tool for automatic file format identification in the ePrints ingest process.

TNA is also working with Andrew Wilson (AHDS) on a proposal (due for submission at the end of June 2006) for the JISC repositories programme call. This work would look at the properties that need to be understood about a file in order to migrate it, and how to measure whether file migrations have been successful.

11.2.2.2 Current status

In full service (not JISC funded work)

There is a need for further development, and the following additional services are planned:

- Characterisation
    - Automatic identification of file format when format is unknown
    - Automatic validation of file format where format is known
    - Property extraction: extracting and measuring those properties of a digital object that must be understood to support preservation and access
- Preservation planning service
    - Risk assessment (to determine when preservation action is needed)
    - Technology watch (to update the risk assessment criteria
    - Impact assessment (to determine the impact of changes in risk on a particular collection)
    - Preservation plan generation (to determine what action to take)
- Preservation action service
    - Deploy the relevant tools to perform the agreed preservation actions

More generic developments, for example resolution mechanisms and a method whereby PUIDs can be used to point directly to information in the PRONOM database, that support how the registry can be accessed, are also planned.

11.2.2.3 External element for JISC IE

Format registry services are needed for the JISC IE, and will be fundamental to the effective operation of repositories over the long term. PRONOM offers a different service to that being developed under the DCC programme: the DCC RepInf Reg/Rep emphasis is on structured scientific data, whereas PRONOM is focused towards office type formats and image, sound and video formats.

The file format identifiers (PUIDs) could potentially be developed into an international standard number.

11.2.2.4 Risk assessment

JISC is not currently developing an equivalent service to PRONOM. Although GDFR might potentially offer an alternative it is still at a very early stage; additionally, it will be a distributed network and PRONOM may act as a GDFR node.

The major players at present are academic or public organisations; it is unlikely that anything will be developed in the commercial sector to fill this gap.

11.2.3 Global Digital Format Registry

The Global Digital Format Registry (GDFR)[70] is being developed at Harvard University. This two-year project, expected to finish in January 2008, will provide sustainable distributed services to store, discover and deliver representation information about digital formats.

Preparatory work included the gathering of use cases from institutional participants. This has enabled the project to determine the categories of use of a format registry.

It is planned to develop the format registry as a distributed registry with a number of nodes, in order to decrease its reliance on any particular institution or funding stream, and to maximise participation. One node would be designated as the root node, responsible for registering immediate child nodes (i.e. top level nodes) and the release of vetted information. The timescale for the project is:

- Theoretical aspects (data model, architectural model, network protocol, editorial process, etc) to be finalised by August 2006.
- First attempt reference implementation due in place by February 2007.
- Root node in GDFR to be fully operational by January 2008.
- Remainder of network (child nodes) due to come online shortly after the root node.

11.2.3.1 What the project aims to deliver

Develop a format registry that will support the following categories of use, either directly within GDFR or as compatible third-party services:

- Core services to be developed within the current project
    - Look up the characteristics of a format
- Aims for the future
    - Identify the format of a file
    - Validate the format of a file
    - Assess the risks associated with a format (e.g. risk of obsolescence)
    - Determine the optimum migration path between original and display formats (delivery)
    - Determine the optimum migration path between original and similarly functional format (transformation)

11.2.3.2 Current status

This is not a JISC funded project and is still at an early stage of development.

11.2.3.3 External element for JISC IE

The National Archives has indicated willingness to be a GDFR node, and the Digital Curation Centre has also expressed interest in contributing. Outside the UK, the Library of Congress has indicated it would be willing to be a GDFR node.

11.2.3.4 Risk assessment

Governance is a big issue and there has been much discussion about who would be responsible for the root node, and what rights they might have with regard to the child nodes. There is a political dimension to this and therefore there is a genuine risk that this could be an insurmountable obstacle to the emergence of a genuinely global network.

11.2.4 VERSIONS

While representation information and file format are crucial to successfully accessing resources, the related issue of variant versions contained in certain databases and repositories should not be ignored. Variant versions arise (a) as the result of collaborative authoring, (b) as part of the publication process and (c) as intended for targeted audiences. Search processes therefore need to be able to identify the version of a resource, and there are implications for shared

infrastructure services such as authentication and authorisation where one version is freely accessible, while another is subscription limited.

The JISC Digital Repositories Programme has funded the VERSIONS (Versions of ePrints – user Requirements Study and Investigations Of the Need for Standards)[71] to investigate attitudes and practices regarding versions of academic papers in digital repositories. The project consortium is led by the London School of Economics and Political Science, with the Nereus Consortium of European economics research libraries as associate partners. The project is taking place between July 2005 and January 2007.

The project has a focus on eprints in the subject discipline of economics and takes a comparative view by drawing on established partnerships and experience with European libraries specialising in economics. A major part of the work is to gather feedback and opinion from key people in the field who have involvement in repository activity from a variety of perspectives.

11.2.4.1 What the project aims to deliver

The project aims are:

- To clarify the position on different versions of academic papers in economics available for deposit in digital repositories, in order to help build trust among academic users of repository content
- To produce a toolkit of guidelines about versions for authors, researchers, librarians and others engaged in maintaining digital repositories
- To propose standards on versions to JISC to inform discussions and negotiations with stakeholders

11.2.4.2 Current status

Exploratory project

11.2.4.3 Support function within JISC IE

The outcomes of this project would provide support for repositories by providing standards and guidelines regarding versions of academic papers. This will be of benefit to ingest procedures and search processes.

11.2.4.4 Risk assessment

Individual repositories will design their ingest procedures with varying procedures for version recording and control. This may lead to a situation where some repositories are more trusted than others because of the quality of their metadata.

11.2.5 Format Conversion Services

Annex A of JIIE(05)36) mentions (at 4.3) the possibility of developing a File Format Migration Service. This would be a structured network service that accepts a resource in an obsolete file format and returns a representation of it in a current file format.

There is no JISC project as such for this type of service, although it could be built on the DCC RI RegRep. PRONOM aims to develop a file migration service as part of its preservation planning and action strategy but has no delivery date set for this. GDFR could also potentially develop such a service, but is at a very early stage.

It should be noted that format conversion is not always simple, and may be inexact. The success of the migration depends on the nature of the source format and the destination format and also the content. Of interest in this connection are: EAST, TOM, FRED, JHOVE; there has been some experimentation in carrying out file format conversions using combinations of FRED, TOM and JHOVE.

EAST[72] is a data description language for scientific data. Using EAST to define a regular data structure (e.g. plotting one variable against another) enables format conversion. There are plans to use EAST in further work on the DCC RI RegRep.

TOM[73] (Typed Object Model) is a system for managing diverse data formats; it is partly registry, partly a set of services. It is being developed at the University of Phildelphia by a team led by John Mark Ockerbloom and demonstrates how it might be possible to identify the right application tools for a specific conversion.

FRED was in effect an early demonstrator of the principles of GDFR using TOM. It contains details of only 5 or 6 formats.

JHOVE (JSTORE/Harvard Object Validation Environment)[74] is a format identification and validation tool, similar to the TNA DROID tool. It can answer questions such as 'is this a valid jpeg file?' and extract metadata from the file.

Elsewhere, the Australian Partnership of Sustainable Repositories[75] (APSR) is developing the AON service for the notification of obsolete formats.

11.2.6 JISC Future Strategy

Note: The DCC RI RegRep and PRONOM registries address different audiences and might, between them, offer the required representation information required for the JISC IE.

- ***Recommendation: At some point before the end of DCC funding, JISC should review the RI RegRep for potential long-term support.***
- ***Recommendation: JISC should talk to The National Archives (TNA) re next steps for PRONOM. There may be potential for TNA to collaborate with other projects.***
- ***Recommendation: JISC support the potential for format conversion work within PRONOM and DCC work rather than seek to develop any new service.***
- ***Recommendation: JISC should keep in touch with Global Digital Format Registry (GDFR) progress.***

11.3 Managing Digital Resources

Introduction by Andy Powell, Eduserv

In the current information landscape, resources are typically discovered through structured meta-searching approaches such as those offered by the JISC Information Environment portals of one kind or another or through less structured full-text indexes such as Google. In the former, discovery is based on the search and retrieval of relatively simple resource metadata such as 'simple' Dublin Core or IEEE LOM. In such cases, an identifier of the licence under which the resource is made available (often a URL) may be added to the metadata record. In the case of resources discovered through Google, the licence may be linked to the resource informally (for example, by including a licence icon somewhere on the page), formally (for example by using the HTML <link> tag) or in some cases not at all.

However, even in cases where the identifier of the licence under which the end-user can use the resource is made available, that may not be sufficient because:

- Where the licence identifier is not a URL, it will not be easy find out more about the licence.
- Where the licence identifier is a URL, there are no guarantees as to what kind of information will be provided at that URL. In some cases, the URL will resolve to a machine-readable licence, in others to a short description of the licence, in others still to a full legal document.

A licence registry shared service component would allow end-users and other services to 'look up' the licence (based on its identifier) and be supplied with a consistent set of machine-readable information about the licence. The intention would be to support two key functions:

- providing a view on what licences are in use (i.e., who is doing what)
- providing a persistent record of what was licensed, under what conditions, and when.

11.3.1 SHERPA-ROMEO

The Rights Metadata for Open archiving project (ROMEO)[76] was a JISC funded project carried out during 2002 – 2003 at the University of Loughborough. It investigated the rights issues surrounding the self-archiving of research in the UK community under the Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH). By surveying the academic community, it ascertained how give-away research literature and metadata was used and how it should be protected.

The ROMEO Project aimed to generate some simple rights metadata by which academics might describe the rights status of their open-access research papers (ePrints) and also to provide a means by which OAI Data and Service Providers might assert the rights status of their metadata under the OAI-PMH. The project identified a potential solution as to how these rights might be disclosed under the OAI-PMH.

Following negotiations with the ROMEO Project team and JISC, another project, Securing a Hybrid Environment for Research Preservation and Access (SHERPA), developed the outcomes of ROMEO into a database-driven searchable service and knowledge bank of information. SHERPA is currently working on two other projects.

PROSPERO is a 3-year project to develop an interim repository that can be used for self-archiving when the author's institution does not have its own repository. Papers would be archived for up to 5 years, and transferred to the appropriate institutional repository when this becomes available. Within this project, there could be some scope to develop ROMEO searching to provide further options.

SHERPA is also working with the Wellcome Trust, who intend to establish an archive for the output of projects and research that they fund. On their behalf, SHERPA is exploring whether publishers will allow such deposit. It is hoped that some of this information can also be added to the ROMEO database.

Funding has been found for a temporary post (15 months from early July 2006). The post holder will work on PROSPERO, Wellcome Trust publisher audit and ROMEO. The work will be evaluated after 6 months and continuation funding may be looked for.

11.3.1.1 What the project aims to deliver

- Provide and maintain a Web accessible and searchable database that records a selection of publisher's copyright transfer agreements for author self-archiving, using categorised agreements.
- Use different colours to highlight publisher's archiving policies, differentiating between four categories of archiving rights.
- Continue to extend the dataset upon which it is based through updates and appropriate suggestions from the user community.

11.3.1.2 Current status

In full service

ROMEO provides only a human-readable interface enabling users to search the database for specific publishers or browse the complete list. For each publisher a series of facts are listed, and a ROMEO publisher category (white, yellow, blue or green) is assigned. Further search options and results displays will be developed as part of the work on PROSPERO.

The database is not comprehensive and there is no funding for systematically extending its coverage. However, some additional entries are expected to be added as part of the Wellcome Trust work.

ROMEO is well used outside the UK and the viability of establishing 'franchise offices' of trusted editors, who would be able to edit and add entries directly, is being explored. This would expand the database substantially. Current contacts are Australian Research Repositories Online to the World (ARROW)[77] in Australia, Vanessa Proudman at the University of Tilburg (contacting 200 Dutch publishers and economics publishers), Lund University (to cover Scandinavian, Baltic and

Slavonic publishers), Nottingham University campuses in China and Malaysia, and other potential contacts in Germany and the US.

11.3.1.3 Support function within JISC IE

This resource provides academics wishing to deposit papers in institutional and subject-based repositories with a single source of information about the archiving policies of some publishers. There is potential for an M2M version of this service to be used by institutional and subject-based repositories, such that when someone wishes to deposit a paper, the system can alert them to any problems or restrictions such as embargo periods.

Ideally such an M2M version would be comprehensive in coverage, which is not currently the case. Re-using information collected for the Wellcome Trust work will help as will the plan to recruit trusted editors world-wide. However, ROMEO would benefit from additional funding in the short term to add more data and longer term funding to maintain the service.

There is potential for collaboration between SHERPA/ROMEO and the ONIX for Licencing Terms work; opportunities for collaboration should be more easily identified as a SHERPA/ROMEO staff member now takes part in the monthly international teleconferences.

11.3.1.4 Risk assessment

The current human-readable interface is useful, but the JISC IE would benefit from a service with M2M capability. Current ROMEO development is being 'piggybacked' on to other work, which means that development is episodic and may terminate at some point before reaching its full potential.

11.3.2 ONIX for Licensing Terms[78]

ONline Information Exchange (ONIX) is a family of XML formats for rich metadata communications formats that have been developed by Book Industry Communication (BIC)[79] / EDItEUR[80]. The first ONIX format was the ONIX for Book Product Information Message, the international standard for representing and communicating book industry product information in electronic form. This was followed by ONIX for Serials, a family of XML formats for communicating information about serial products and subscription information.

A new format, ONIX for Licensing Terms, is currently being developed to address the need to express licensing terms for digital resources in a standard Extensible Markup Langauge (XML) format, link them to digital resources and communicate them to users. It will be a family of formats; the first of these to be developed is the Publisher Licence Message, which is focused on the needs of libraries. A potential further message format would be a publisher search engine message, since publishers are now – galvanised by recent Google projects – starting to build their own repositories and will need some machine-readable way of expressing usage terms to search engines.

11.3.2.1 What the project aims to deliver

ONIX for Licensing Terms aims to deliver licence terms digitally, allowing vendor and subscriber to communicate effectively. The format is built on the functionality described in the Digital Library Federation (DLF)[81] Electronic Resource Management Initiative (ERMI)[82] project, using the <indecs>-based rights model, which views licensing conditions as events – permitted, prohibited, required, etc.

The Publisher Licence Message is a structured expression of a publisher licence for the use of (digital) resources from the publisher to the agent or subscribing institution. The Licence itself will still be paper-based but the message will contain metadata on 'actionable parts' (permitted/prohibited usages and related conditions, notice periods and permitted date changes, and bases of fee calculations) and will quote non-actionable information. A draft version was published in February 2006.

The format was the subject of a JISC /Publishers Licensing Society proof of concept project in 2005, and two recent JISC projects. The first mapped the Wiley science licence into ONIX and is a collaboration between BIC, John Wiley and Cranfield University; the report is available at the EDItEUR site[83]. The second is an exploration of the tools and services required by publishers to enable them to create the messages; [this collaboration between BIC, JISC and

the Association of Learned and Professional Society Publishers (ALPSP)[84] is due to complete at the end of June 2006.]

At a joint BL/JISC Workshop on Digital Rights Management, held on 24th April 2006, a suggestion was made that JISC carry out a trial to put JISC model licences into ONIX for Licensing Terms format. The authors understand that JISC is now pursuing this via the JISC PALS Metadata Interoperability projects[85].

11.3.2.2 Current status

Project with identified service potential (not JISC funded)

This is currently at the demonstrator stage, but BIC would hope for a fairly rapid transition to use. At present the specifications are there but not the tools and the use of such messages needs to be piloted.

11.3.2.3 External element for JISC IE

As the number of digital resources in library collections grows, libraries have increasing difficulty complying with the widely differing licence terms applied to resources by their creators and publishers, and communicating these terms to users. Although cataloguing formats such as MARC 21 can already hold some of this type of information, it may not be sufficient and may require display customisation to present together information held in different fields. Individual libraries will also need to key in such copy specific information into the records. ONIX for Licensing Terms offers the potential to provide appropriate information along with the supply of the resource.

Within the JISC IE, the ONIX for Licensing Terms messages could potentially be used in repositories, JORUM and expressions of Creative Commons Licences.

The British Library is interested in using ONIX for Licensing Terms for legal deposit and document supply processes.

11.3.2.4 Risk assessment

JISC is currently not developing anything similar and ONIX for Licensing Terms, which therefore fills a gap, is being developed internationally. However, tools still need to be built. Publishers will not, and libraries usually cannot, pay staff to manually input licence details each time, so a tool that allows them to easily and quickly choose clauses from a list of options is needed. In addition to the possibility that such licences could be supplied with resources, it would also enable libraries to record details of licences for resources they already have; typically at present, such paper licences are filed somewhere but no details are added to catalogue records.

11.3.3 JISC Future Strategy

- ***Recommendation: Carry out a small-scale study to examine synergies between ROMEO and ONIX for Licensing Terms.***
- ***Recommendation: Carry out an evaluation for funding ROMEO development to M2M capacity.***
- ***Recommendation: JISC should maintain contact with BIC and seek collaboration opportunities (tool building and piloting) in the continuing development of ONIX for Licensing Terms.***
- ***Recommendation: JISC should investigate the need for a licence registry shared service component, initially by funding a pilot project.***

11.4 Service Registries and supporting metadata

Portals can provide a route for users to access to the collections they are entitled to use, but need metadata about these collections in order to provide the appropriate access. Portals, brokers and aggregators also need this metadata to determine how to interact with available network services. Service registries and collection description metadata support these requirements.

11.4.1 Information Environment Service Registry (IESR)

The Information Environment Service Registry[86] aims to create a reliable source of information that other applications, such as portals, can freely access through machine-to-machine protocols, in order to help their end users discover useful resources.

The IESR contains information about collections of resources themselves, technical details about how to access the resources, and contact details for the resource providers. The collection information is based on the RSLP (Research Support Libraries Programme) Collection Description Metadata Schema.

The current project phase ends in July 2006. In May 2006, funding was agreed by JISC for a further three years. The plan is to move from a project to a 'service-in-development'.

The full scope of the registry has yet to be fully defined but in the forthcoming phase content will be expanded to encompass non-JISC service providers. This will include institutional resources (e-print repositories, OPACs), and resources of other Common Information Environment partners.

Initial research amongst the stakeholder community indicated that 83% of stakeholders would make use of the IESR[87]. It was recognised that a major challenge for the project team was to convert that theoretical willingness into real use of the registry.

The registry is held in an XML database using Cheshire information retrieval software. Full IESR metadata records are available in XML, via Z39.50, OAI-PMH, and an OpenURL Link-To Resolver. Every entity registered in IESR is assigned a unique global identifier, using the PURL-based Object Identifier (POI) scheme. IESR identifiers are resolvable via PURL/POI and the IESR OpenURL Link-To Resolver.

11.4.1.1 What does the project aim to deliver?

- A fully functional demonstrator providing a central source of information about electronic resources (within the JISC IE and beyond) and the ways in which they can be accessed
- Interfaces according to several standard protocols: Z39.50, OAI-PMH, OpenURL Link-To Resolver, Web Search (for human-readable access), SRU and UDDI.

11.4.1.2 Current status/development

Project with identified service potential

The project recognises the need to add new content, and also to update existing content. An IESR workshop (*Include Electronic Services in a Registry*) was held in January 2006. Participants showed interest in contributing data, and new contributions are now starting to be received.

A dilemma exists, since in order to prove its usefulness and secure support (in terms of data input), IESR needs a critical mass of data to demonstrate a fully functional service. The project needs portals and other services to use the registry in a serious way to show that it is useful.

The project plans to develop use cases. These should help to demonstrate usefulness.

IESR have a Dewey licence, and plan to include a Dewey term for every record.

OpenURL resolvers are not currently included in IESR, although this is a desirable development. Either libraries would need to be encouraged to supply data, or it could be harvested from the OpenURL Router.

The next phase includes a workpackage looking at terminologies, and may involve liaison with the HILT project (see section 11.7.2)

IESR have collaborated with the Ockham Initiative[88] which is using the IESR metadata schema for the development of its registry.

There has been a lot of discussion about distributed registries versus centralised, partly in the context of Ockham. A workshop was held during 2005 in Warwick which included e-Science and US participants. Despite the interest shown at the workshop, issues raised regarding distributed registries have not yet been taken forward.

IESR is also likely to have some collaboration in the next phase with the eScience community. They use Grimoires[89] registry software developed at the University of Southampton.

IESR has been liaising with the e-Framework development activity. At the time of writing discussion is ongoing about whether the e-Framework should have a registry of services. DEST has used the IESR metadata schema for related e-Framework development in Australia.

Collaboration is planned with the OpenDOAR[90] directory of open access repositories development at the University of Nottingham and Lund University in Sweden. OpenDOAR plan to include an M2M service offering a lookup service over several hundred open access repositories.

IESR have a Creative Commons licence. There could be licence issues in the future if records are shared between registries internationally where different policies apply.

11.4.1.3 Support function within JISC IE

IESR forms part of the basic shared services infrastructure, directing portals and other applications requiring M2M communication to relevant resources.

However, given that a registry is an 'invisible' service, the community is only aware of it when it is not working.

11.4.1.4 Risk assessment

There is currently no other M2M service with the potential of enabling portals to easily find out about the full range of relevant resources. If IESR were not available, the JISC IE may not be used to its full potential, since unknown resources may not be discovered. Without a registry service, portals would need to develop individual M2M links.

There are several commercial registry services available, but those from library vendors (e.g. Talis, Ex Libris) concentrate on bibliographic services, whereas IESR covers all types of resources.

11.4.2 Collection Description Metadata

Collection Description Focus (CDF) was originally a jointly funded activity of the JISC, MLA and the British Library, and it is now part of the core funded work programme at UKOLN. It was set up to provide support to projects implementing collection level descriptions and the wider community, and consequently much of the work has involved workshops, presentations, briefing papers and case studies, and providing advice to individual projects.

However, a key component in the work of the Focus is the development of metadata standards for collection level description. Michael Heaney's report *An Analytical Model of Collections and their Catalogues*[91] defined the entity-relationship model that provides the structure for the metadata standards. Initially two standards were based on this model, the RSLP metadata schema and the SCONE metadata schema.

The RSLP collection-level description metadata schema was developed for the RSLP Programme and a number of collection description databases have been implemented using this schema, with the result that it has become a de facto standard. Although theoretically possible to cross-search these databases, in practice no such attempt has been made. Implementation experiences have identified some gaps in the RSLP schema and suggestions made that the schema should be revised and updated; the revised schema could then be candidated for formal adoption as a national and/or international standard. The IESR metadata schema uses the RSLP schema for its collection elements.

The RSLP schema has also been used as the basis of MLA's Corncopia database of UK cultural and heritage physical collections and the European MICHAEL database of physical and digital cultural and heritage collections; however, there has been some divergence from the schema, which may affect interoperability. Outside the UK, the RSLP schema (with some modification) has also been used in the creation of a collection registry for digital collections at the Institute of Museum and Library Services (IMLS)[92] at the University of Illinois at Urbana-Champaign.

The SCONE collection-level description metadata schema was developed for the SCONE database; this implements the Heaney model in more detail than the RSLP schema, and theoretically, since they are both based on the same model, cross-searching SCONE and RSLP based databases should be possible.

Subsequently, the Dublin Core (DC) collection description community decided to develop a Dublin Core Collection Description Application Profile (DC CD AP). This defines only those attributes which describe the collection itself; other attributes in the Heaney model are defined by other DC elements (e.g. Agents).

The NISO Metasearch Initiative is also working on a Collection Description Specification[93]. This is currently a standard for trial use; it is probable that this will be awarded formal standard status at some point in the future.

11.4.2.1 What the RSLP Collection-level Description Metadata Schema aimed to deliver

The metadata schema was designed to support collection description information required for the RSLP digitisation and retrospective cataloguing programme strands, although it was deliberately designed with wider use in mind and intended to cover all types of collection, both physical and digital, in all domains (museums, libraries and archives) and all sectors (education, cultural heritage, etc.). There were two deliverables:

- A metadata schema for collection-level description
- An empty Access database using the metadata schema that projects could download and populate with data about their collections.

11.4.2.2 Current status of RSLP schema

In full service

However, a number of implementations have identified a need for revision.

11.4.2.3 What the Dublin Core Collection Description Application Profile aims to deliver

The RSLP schema uses a number of existing DC elements and attributes but neither core nor qualified DC were sufficient to describe all attributes of collections, so further attributes were defined. The RSLP schema was then used as the starting point for a DC CD AP.

11.4.2.4 Current status of DC CD AP

Work on the application profile is almost complete and the intention is to present the final sections to the DC Usage Board in October 2006 for approval.

11.4.2.5 Support function within JISC IE

The RSLP schema (and any subsequent revised versions), the SCONE schema and the DC CD AP will all support resource discovery at the collection level.

11.4.2.6 Risk assessment

Although there have been few recent implementations of collection description databases, this is in large part due to lack of available funding and funding initiatives. Since many of these implementations have used the RSLP Collection Description Metadata Schema, this schema has become a 'de facto' standard. However, the implementation-specific extensions to the schema, which have also fed into the development of the DC CD AP, mean that there is a good case for revising the schema and registering it as a formal standard. A complicating factor is that the one area where collection description is being actively taken forward is the MLA Cornucopia and European MICHAEL databases; their metadata schemas are based on the RSLP schema but with implementation-specific variation. There is a need to support interoperability by taking the RSLP schema forward as a formal standard as it will be better to map all variants to the standard than to make multiple mappings of variant to variant.

11.4.3 JISC Future Strategy

- ***Recommendation: it would be useful for JISC to offer some guidance on IESR content scope, in terms of coverage (eg JISC IE, UK-wide, European, international), and resource type (repositories etc).***

- ***Recommendation: resourcing for populating the IESR database should be considered.***
- ***Recommendation: further discussion is needed about the issue of distributed IESR registries.***
- ***Recommendation: collaboration should be pursued with OpenDOAR in order to ensure compatibility with IESR.***
- ***Recommendation: a realistic 'business plan' should be developed for IESR; it is unlikely that service registries can be sustainable without some form of support.***
- ***Recommendation: JISC should continue to fund work to develop and maintain a standard for collection-level description metadata.***

11.5 Metadata Schema Registries

A metadata schema registry is an application that provides services based on information about metadata vocabularies, the component terms that make up those vocabularies, and the relationships between terms.[94]

'A metadata registry provides machine-readable information about the metadata schemas in use by particular metadata-based services. The primary intention of this service is to allow portals, brokers and aggregators to automatically determine information about appropriate search terms and the structure of metadata records that will be returned to them. However, metadata registries also provide a useful human-oriented service, allowing people to see what metadata schemas are in use by which services - providing a basis for metadata schema sharing and re-use.'[95]

The human facing web service is therefore of benefit in its own right – a schema registry is not just an M2M service.

11.5.1 IEMSR

The JISC IE Metadata Schema Registry[96] project aims to be the primary source of information about particular metadata schemas and application profiles recommended by the JISC IE Standards Framework. Phase 2 of the project runs from July 2005 until September 2006.

There are three functional components:

- Registry Data Server - an RDF application providing a persistent data store and APIs for uploading data (application profiles) to the data store and for querying its content
- Data Creation Tool - supports the creation of RDF Data Sources (application-specific profiles) for use by the Registry Data Server
- User Web Site Server - allows a human user to browse and query the data (terms and application profiles) that are made available by the IEMSR Registry Data Server

The Registry is targeted at the UK education community where both Dublin Core (DC) and IEEE Learning Object Metadata (LOM) standards are used. IEMSR focuses on both DC and IEE LOM application profiles. However covering two different schemas has added a layer of complexity to the project since the two standards have incompatible data models, so designing user interfaces for an integrated tool is complex. One option in future may be to split the registry into separate services for different metadata standards, whilst presenting the human user with an integrated web interface.

There has been some feedback to encourage expanding the Registry to cover other metadata standards in use within the JISC IE (e.g. METS). However inclusion of additional standards may require integrating yet another data model so more flexibility would be required. The current registry software was developed primarily with the Dublin Core data model in mind – adding additional models increases complexity.

The Registry could also act as a maintenance agency for application profiles.

Project staff have been demonstrating use scenarios to stakeholders such as JISC and the British Library during June 2006. The British Library would like to implement the registry software to manage their own application profiles in use within the BL.

11.5.1.1 What does the project aim to deliver?

- A pilot registry service providing information about DC and IEEE LOM metadata schemas and application profiles.
- Enhanced registry software – develop web interface to registry, registry server, schema creation tool
- Use scenarios for the registry – these have helped to specify priorities for the development of IEMSR software components and to target dissemination activity. This work also defines more clearly the benefits that would be delivered by a pilot registry service
- A business plan, providing an outline proposition and marketing plan
- Liaison with the JISC Standards Catalogue activity to ensure that the information provided is consistent with the data held/provided by IEMSR
- Liaison with the JISC/DEST e-Framework activity to ensure that the outcomes of the IEMSR project regarding the functions of a metadata schema registry are integrated with the Framework.

11.5.1.2 Current status

Project with identified service potential.

There has been exploration of inter-working between the between the IESR and IEMSR. IESR would offer a third party agent (e.g. portal) information on which metadata application profile a service used, and the IEMSR would provide information on subject schemes in use within that profile.

IEMSR is already liaising with a wide range of organisations and related initiatives, including the British Library, Becta, The European Library[97], the National Science Digital Library (NSDL) Metadata Registry[98], DCMI Registry and the Dataset Acquisition, Accessibility and Annotation e-research Technology project (DART)[99] Metadata Schema Registry[100] in Australia.

The project is interested in collaborating with the e-Framework on registering application profiles.

11.5.1.3 Support function within JISC IE

A registry service is a basic middleware component for metadata management.

11.5.1.4 Risk assessment

If the IEMSR did not exist there would be no single point of access for information about application profiles and therefore effort would be duplicated; interoperability would be compromised since services would be encouraged to develop their own application profiles rather than re-use existing profiles. Communication between services generally would be more difficult.

11.5.2 JISC Future Strategy

- ***Recommendation: JISC may need to consider whether separate services are required to manage DC, LOM (and other future) schemas.***
- ***Recommendation: JISC should consider how to encourage the IEMSR registry to be populated in order to achieve a critical mass of data.***
- ***Recommendation: A 'collection policy' for inclusion of schemas needs to be agreed.***

11.6 Institutional Profiling Services

11.6.1 Institutional Profiling and Terms & Conditions Services Scoping Study

An institutional profiling service is used by resolvers to discover information on institutional services and preferences for OpenURL resolution. A terms and conditions service provides machine-readable information about rights held in resources. These were identified in the Shared Services Development Plan but had not been fully addressed.

The study was carried out by EDINA in 2004[101]. It concluded that an institutional profiling service would be desirable to improve communication with the members and services of institutions and the effectiveness of the JISC Information Environment. It recommended a distributed service, best achieved via a pro-forma that institutions could populate locally for M2M processing by other services for specific purposes.

11.6.1.1 What did the project aim to deliver?

The project aimed to re-examine the roles of institutional profiling and terms and conditions services within the shared services model, and formulate concrete proposals for their practical implementation.

The report set out the data that would be required for an institutional profiling service, the sources of that data, rights management issues, service development criteria and service proposals.

11.6.2 JISC Future Strategy

The approach the JISC Executive recommended was that the immediately feasible developments in the area of institutional profiling were small-scale in nature and should be pursued through existing initiatives such as IESR, the OpenURL Router Service or the WAYF service. On the other hand, there may be reluctance within existing projects to widen their scope[102].

IESR did consider this report and added some hooks into the metadata in case it was decided this should go into IESR[103].

- ***Recommendation: JISC should review the approach recommended by the EDINA study and discuss practical implementation with IESR and other relevant projects.***

11.7 Terminology Services

As noted elsewhere, UKOLN is also carrying out a Terminologies Review on behalf of JISC, due to report in July 2006. This will include a more in-depth analysis of JISC and other terminology services.

11.7.1 GeoCrossWalk[104]

Existing resources within JISC IE do not currently exploit geographical searching of resources in any meaningful or consistent way. Different types of geographical reference or coding conventions are employed by different services and the majority of these have only minimal (if any at all) geographic indexing of resources. Most of the resources have some form of implicit geographic reference, such as place name, county name, postcode, etc. but there is no common agreed referencing type that is used by all and simple mappings are not always possible. Postcode boundaries do not match electoral boundaries, administrative areas change over time, and some features map to multiple instances of specific coding conventions – for example, a city will be represented by a large number of postcodes.

Phase 1 assessed the feasibility of developing and providing an online, Z39.50 compliant, fast, scaleable and extensible British and Irish gazetteer service. Phase 2 created a demonstrator gazetteer service; Phase 3 then developed a functioning, scaleable gazetteer service suitable for integration into the JISC IE as a shared service. Phase 4 is concentrating on a middleware infrastructure and developing and maturing a variety of business case scenarios for the

sustainability and further development of the service. (The project name was originally named GeoXwalk).

11.7.1.1 What the project aims to deliver

GeoCrossWalk will provide a mechanism by which one geography coding can be 'crosswalked' into another representation. This will be achieved by creating a database of geographical features (such as towns, rivers, woodlands and counties), their name and location – in other words, a gazetteer. A feature's location is not stored as a point but rather as a "footprint"; for example, places are stored as polygonal footprints, and rivers as linear footprints. In this way, for example, place names can be linked to postcodes, and counties to parishes. Furthermore, it would mean that spatially complex queries could also be resolved, such as "which rivers are near Banbury?" This facility would (a) enable services utilising GeoCrossWalk to be agnostic about which geography referencing their users deploy to search resources and (b) offer a much richer potential query environment to end users. A part of the project developed a 'geoparser' which can spot place-name references in text and look them up against a gazetteer (GeoCrossWalk). The parser allows textual geographic references to be turned into numeric references (co-ordinates such as latitude/longitude) and those in turn can be 'crosswalked' by GeoCrossWalk into other geographies, such that a news item on a website that refers to Leicester could be searched by e.g. postcode. Specific deliverables are:

Phase 3. Functioning, scaleable gazetteer service suitable for integration into the JISC as a shared service: deliverable achieved.

Phase 4: Business plan detailing various sustainability and future development scenarios; a quality assured GeoCrossWalk database and production-level service; showcase examples of GeoCrossWalk-enhanced JISC service(s); and marketing and publicity materials. This work includes identifying JISC services and hosted online facilities, which could benefit from location based searching, and contacting key services to discuss their potential use of GeoCrossWalk; this includes meetings with potential non-JISC users of the service: (the British Geological Survey (BGS), the Royal Commission for Historical and Ancient Monuments Scotland (RCAHMS) and the Office of the Deputy Prime Minister). Due to finish 31st July 2006.

11.7.1.2 Current status

Ready for consideration for transition to service

Phase 4 contacts have identified more interest from non-JISC agencies than JISC projects, as the external agencies are more mature. However, taking this interest further is limited by two factors. Firstly, there are licensing issues (e.g. for Ordnance Survey date) that still need to be resolved; and secondly, agencies are discouraged when they learn that GeoCrossWalk is still a project and not a service with commitment to future maintenance.

Within JISC, GeoCrossWalk have carried out some pilot testing work with AHDS.

Implementing GeoCrossWalk for any service will require some changes to provide a client interface that will successfully integrate with that service. It is likely that such client interfaces will need to be developed for each service. However, this should not be a difficult task and could be achieved through short-term consultancy work.

If GeoCrosswalk were to be moved to production level, a technical analysis would be required to ensure that the platform could support large scale use.

Although JISC has not restricted GeoCrossWalk to call funding, the project feels development has not been continuous but has taken place in spurts. Initial assumptions were that JISC would fund the transition into a service but the process has not been as straightforward as the project team would have liked. In the latest phase they have been asked to come up with a business model, albeit with little guidance, they feel, on the metrics to gauge the value of projects to the JISC IE. As part of the business model work, potential 'customers' outside HE have been contacted but marketing middleware (always difficult as it is an invisible product) has been made harder as what is offered still has project status. The JISC Executive recognises that this is an issue and has been actively working on a process that will help in terms of transitioning projects to service. There have been discussions with the Sub-committees primarily involved in this process and also there was an away day dedicated to working out how to proceed with the

JISC Development Group and the Services and Collections Teams. The process is progressing and is being tested at this current time.

11.7.1.3 Support function within JISC IE

GeoCrossWalk is potentially of use to any service requiring geographic searching. It would:

- enable users to cross-search various resources using different geographic referencing conventions
- enable users to use their preferred geographic referencing convention for searching, irrespective of whether the targeted resources used that same convention.
- enable users to move from a place-name reference in a text to a further geo-referenced search.

11.7.1.4 Risk assessment

If GeoCrossWalk is not approved for transition to service, the initial impact of this will be limited, as the current situation would remain. As time passes, it will become more likely that external agencies – perhaps commercial agencies – will develop a similar product as the need will still be there. If this happened, then JISC could face the situation where they will need to at least support integration with such a product, and that any such product might only be available on a commercial basis.

There is at present no exact equivalent work going on elsewhere. The Alexandra Digital Gazeteer and the Digital Library Project at the University of Berkeley have demonstrated the use of georeferencing gazetteers to provide *indirect* georeferencing to geospatial datasets. Gazetteer Protocol uses a different approach to geo-referencing, as do Google Earth and Google Maps. It would be useful for the technical evaluation to describe the different approaches.

11.7.2 HILT[105]

End-users face difficulties when they want to carry out a subject search or browse across a number of resources that are indexed using different controlled vocabularies. A variety of controlled vocabularies exist to meet the requirements of specific communities of use and will continue to be used but there is an overarching requirement to provide a subject-based facility for search and browse across the boundaries of discipline and institution. The HILT project aimed to address these difficulties by developing such a subject-based service.

Phase 1, which reported in December 2001, was a desk-based study that reviewed approaches to improving cross-searching and cross-browsing by subject. It concluded that there was consensus across communities to take work forward through a pilot mapping service.

Phase 2 was funded as a short pilot running from mid-2002 to late 2003. Deliverables were to provide terminology services at the collection level, whilst recognising the need to extend this in the future to item level retrieval. The pilot web user interface used Wordmap; this is a commercially available product that supports management of multiple controlled vocabularies in a single user interface, management of partial views of controlled vocabularies and mapping between different controlled vocabularies. DDC numbers were used as a central spine; these combined with LCSH and UNESCO Thesaurus form the core of the server. Other schemas (MeSH, AAT, etc.) could then be added at a later stage.

At this point the project was reviewed for its potential development as a JISC service.[106] Within the definition of the JISC IE, HILT would be defined as a ‘transactional network service’. As such, according to the JISC IE, HILT would need to be accessed using either the SOAP protocol, or using HTTP GET/POST. But to be delivered in an m2m way, HILT also needs to be accessed in a structured way; for example, using SOAP wrappers around structured query semantics. In addition to the resource itself, the controlled vocabularies and mappings all need to be structured. Though preferable for HILT to follow widely agreed standards, since at that time there were no such widely agreed standards, it was conceded that HILT might best rely on a proprietary product.

A short study on the feasibility of developing SOAP based interfaces between JISC IE services and Wordmap APIs and a non-Wordmap version of the phase 2 demonstrator was carried out between January and March 2005. Conclusions from this study formed the basis of the proposal for phase 3.

Phase 3 began in November 2005 and will focus on creating an m2m demonstrator that will offer web services access. It will run until January 2007. Like phase 2, the demonstrator will be based on the centralised approach to the provision of mapping services, but designed so that a future move towards a more distributed model may be possible.

11.7.2.1 What does the project aim to deliver?

The aim in HILT Phase III is to build an m2m version of the pilot demonstrator service built in Phase II, which will demonstrate m2m terminology services for the JISC IE.

Specific objectives are to:

- Offer web services access via the (SOAP-based) SRW protocol, but design the pilot so that a possible extension offering other protocols (eg Z39.50 or SRU) could be an option at a later date.
- Use SKOS-Core as the mark-up for sending out terminology sets and classification data responses but design the pilot so that adding other formats such as MARC and Zthes would be an option at a later date.
- Provide the pilot datasets, mappings, and functionality capable of servicing the five use cases agreed in the HILT m2m Feasibility Study[107].
- Base the pilot on the centralised approach to the provision of mapping services piloted in HILT Phase II, but design it so that the possibility of a future move towards a more distributed model is kept open.
- Include a pilot implementation of the SRW EXPLAIN function.

Instead of Wordmap, the service will use a 'simpler-to-use' and 'work on SQL server' clone. Users will not access HILT directly as in phase 2, but will use browsers to access services such as GoGeo! and BIOME and the services will interact with HILT via service-based SRW clients and a HILT SRW server. It will therefore allow various JISC services to provide terminology mapping services to their users in a transparent way.

At a broader level, a planned outcome is a better understanding of the working requirements of terminology services, whether centralised or distributed.

An important aspect of the development is its real world application. The SRW clients could therefore be adapted and used within other services as embedded clients. EDINA is developing the GoGeo! Service SRW Client.

There are two aspects to the pilot. The first addresses mapping between schemes to provide interoperable subject cross-searches. A related part of it uses DDC and a collections database to identify collections relevant to a user's subject search. Mapping is only being carried out currently in order to illustrate functionality. A full-scale mapping programme would clearly be necessary for an fully operational service. This could be done in a phased way, focusing on real problems first (eg subject interoperability problems across RDN hubs or ex-hubs) and could piggy-back on a service to supply non-mapping based terminologies information (eg broader and narrower terms from specific schemes such as UNESCO, AAT or IPSV etc).

The second part of the pilot arose out of the use cases. It deals with serving information on stand-alone terminology sets, such as broader and narrower terms, synonyms, scope notes etc. The service is therefore used to enrich users' search terms (although invisibly to the user). Planned schemes to be served include UNESCO, DDC, LCSH, IPSV, JACS, AAT and others. This could potentially become a service after a further transition to service phase and could provide the basis for ongoing mapping work to improve interoperability. Exploratory work on dealing with spelling mistakes, typos etc is also being carried out.

Therefore both aspects of the current HILT pilot could in theory be taken to the 'transition to service' stage, although the mapping side would have to develop gradually over a period of time. The decision on transition to service will be made by JISC (see discussion in section 13.2).

In-depth user testing was recommended by the project at the end of Phase II, but JISC took the decision to move on to m2m development first.

11.7.2.2 Current status

Exploratory or Project with identified service potential

The project is currently funded (to January 2007) to develop a demonstrator service based on current innovative technology.

In the future it would be useful to investigate the implementation of a distributed approach. Given a distributed set of terminologies services there would be potential for mapping to be shared out internationally. It would be useful to explore collaborative work with the Becta Vocabulary Studio in this context.

There is potential for repositories to use HILT services.

Mapping between schemes is often carried out to fulfill local needs, but there are no other known initiatives tackling the issues at a generic level like HILT.

HILT is hoping to collaborate with IESR to select collections. IESR plans to have SRW capability before the current phase of HILT ends, so if this is straightforward, it will be implemented.

A future phase of HILT could look at how folksonomies might be used to improve mapping between user vocabularies and controlled schemes and vice versa.

11.7.2.3 Support function within JISC IE

HILT provides mapping between subject schemes to provide interoperable subject cross-searches; it also provides m2m information about terminology sets which is used to enhance the precision of subject searches.

11.7.2.4 Risk assessment

Cross searching by subject in the distributed, multi-scheme JISC IE, will continue to frustrate users.

11.7.3 Becta Vocabulary Studio[108]

Even within a single subject area such as education, there are often many vocabularies in use and problems arise with these multiple vocabularies, some of which exist in a number of versions. Becta identified a need for a terminologies server type tool/service, to provide an integrated solution that could be used in systems and applications (e.g. learning platforms), learner information (e.g. e-portfolio), assessment and reporting (e.g. e-assessment), learning design, content discovery, harvesting and embedding, and MIS and data discovery. A key requirement was a product that could map equivalent terms in different vocabularies via a common spine, such that any vocabulary could be mapped to any other vocabulary in just two steps.

The Studio is available, free of charge, under licence to agencies involved in the UK educational sector.

Becta have been in contact with English Heritage and MLA regarding potential use by these bodies. For example: MLA would have a need to store the MLA internal terminology being created by TFPL; MLA project Cornucopia uses UKAT, and the EU project MICHAEL uses the UNESCO Thesaurus; local government agencies are required to use the IPSV. There some concern over the storage and maintenance of these vocabularies. At an exploratory meeting in September 2005, there was general interest in the service but several issues were raised.

- The concept spine is in English. It was noted that this could pose difficulties if used for EU projects, since EU policy is that no language has more prominence than another. Even within the UK, there could be difficulties with Scotland and Wales not wanting to use an English spine. The developers thought that this could be avoided by the spine concepts being replaced by codes, which can then be linked to terms in different languages.

- There are problems with degrees of equivalence when mapping vocabularies to one another. For example, 'crown property' exists in English and Spanish law, but not in that of other countries. There are five levels of equivalence: no equivalence, partial equivalence, exact equivalence, inexact equivalence and single-to-multiple term equivalence.
- The spine of concepts started with an educational bias. While initially this meant that certain areas were less well developed than others (with associated implications for some potential users), this drawback is being addressed. The spine is a 'work in progress' and is being added to all the time.

11.7.3.1 What the project aimed to deliver

The Vocabulary Studio was designed as a managed environment in which vocabulary managers and editors can store, edit and maintain their specific vocabulary, mapping their terms to a central spine of concepts. The system offers various levels of permissions for access to different functions and tasks. It also includes a built-in elective system that allows all stakeholders for a vocabulary to participate in decisions to change the vocabulary.

The Vocabulary Bank is a web interface that will enable government agencies to publish the controlled vocabularies that they create and manage within the Studio. These can be used freely in the classification and tagging of educational content for the UK sector. Users will be able to browse and select vocabularies and to download these in ZThes XML format (the accepted standard for thesaurus interoperability) for importing into tagging tools, such as the Becta Tagging Tool. In May 2006, the available vocabularies listed were: Cross Curricular Skills, National Curriculum Programme of Study, National Curriculum Specifiers, QCA Schemes of Work, ACLearn and NLN. There are plans to add GCSE vocabularies in the future.

11.7.3.2 Current status

Advanced prototype service (non-JISC funded)

The advanced prototype service was launched in March 2006 but not publicised while key agencies were formally signed up as users. There are also some queries and issues about the National Curriculum vocabularies, which are currently being resolved, and the live publishing link between Studio and bank is still being developed. The aim is to move to a full service in September 2006.

The service is at present based on human-readable interfaces for vocabulary storage and maintenance, and the download of specific vocabularies into other applications. It has not developed a M2M interface that could be used by a cross-searching application (e.g. such that a search on 'teeth' also searches on 'dental' and 'orthodontics' in databases that use different terminologies).

Input of vocabularies is the responsibility of the vocabulary owners but Becta is currently considering the level of funding and human resources required to maintain it as a service.

There are no technical limits on scaleability. The current licence for the back-end database has a limit of 200 users at any one time. If successful in its present form, it would be appropriate to look at redeveloping the database as an open source product, so that distributed forms could be made available.

11.7.3.3 External element for JISC IE

The Becta service was not designed to provide an M2M terminology conversion service, so is not an alternative to the HILT project. This could potentially be developed if there were demand from the Becta user community, but would still be reliant on the mapping principle.

However, the service as currently available could potentially be a useful tool for other JISC projects, such as repositories, where there is a need to create and maintain vocabularies. JISC may not be aware of this service.

Given the intention in HILT Phase III to examine the possibilities of a distributed approach to terminology services and inter-scheme mapping, there could be benefits from joint HILT and Becta work.

11.7.3.4 Risk assessment

The Becta Vocabulary Studio, like HILT, it relies on mappings. However, at present it cannot provide an alternative to the HILT M2M service.

Repositories and other content providers may require a managed environment for creating and publishing vocabularies. There appears to be no other similar work in the UK or internationally, apart from commercial companies who offer vocabulary development packages. There is some international interest in the Becta service, particularly in Scandinavia.

11.7.4 Other approaches

In addition to the initiatives detailed above, the possibility of adopting alternative approaches to terminology services has been raised during the course of the review. JISC should consider looking at other options. One possibility is using ontologies to understand how terms relate to one another. Another is to use data/text mining engines, that build associations between terms (vectors) and put together clusters of vectors. As an example RedLightGreen (from RLG/OCLC) has used text mining on a large set of catalogue records; it is possible to search on one term, and retrieve records that do not contain that term but are related. A third area to consider is folksonomies, which allow users to add tags to classify information. An example is Connotea[109]. Folksonomies may have a role in improving user accessibility to controlled languages.

In any investigation of alternatives to mapping, consideration should be given to ability to meet specific community needs; they should be measured objectively and demonstrated to work.

The Terminologies Report will provide further guidance in this area.

11.7.5 JISC Future Strategy

- ***Recommendation: JISC should come to a decision on whether GeoCrossWalk can deliver; if yes, approve transition to service.***
- ***Recommendation: JISC should fund a technical review of GeoCrossWalk as part of the transition to service.***
- ***Recommendation: JISC should consider funding a discrete section of the IE to be fully operational (e.g. by populating databases) in order to demonstrate full functionality; this could potentially build on the IE Testbed approach.***
- ***Recommendation: HILT should undertake in-depth user testing in the context of significant (but contained and context-specific) mapping work.***
- ***Recommendation: It would be useful for HILT and the Becta Vocabulary Studio to explore collaborative work, especially given HILT's plans to examine the possibilities of a distributed approach to terminology services and inter-scheme mapping.***
- ***Recommendation: JISC should investigate further the potential utility of HILT and the BECTa service to the repositories programme.***
- ***Recommendation: JISC should consider alternative approaches to terminology services, including ontologies, text mining and folksonomies.***

11.8 Name Authority

As noted elsewhere, UKOLN is also carrying out a Terminologies Review on behalf of JISC, due to report in July 2006. This will include a more in-depth analysis of JISC and other terminology services.

A name authority record comprises the recognised, authorised or prescribed form of a name, usually supported by sufficient information and sources to ensure reliable recognition and use of such a name[110].

The British Library no longer maintains a (de facto) national file, but instead contributes to the Library of Congress (LC) Name Authority File. Authority records in the LC Name Authority File include an LC control number.

Name authority files may be used by many different services. Since they enable a name (author etc) to be uniquely identified, they help cataloguers to avoid using different versions of a name (for the same person), and allow the name to be presented in the preferred format. It is usual to include dates of birth/death in order to distinguish between similar names; some authority files hold additional biographical information.

However problems can arise when a record creator is presented only with an author's basic name: with no extra information available, how can the creator distinguish one Clare Jones from another already existing in the local database? In this case, consulting a name authority file will not assist. This is the primary reason why many databases end up with several different records for the same name. Therefore an author search may not discover all relevant items.

A separate issue is that even when controlled terms are uniformly used, different services may use different standards, e.g. [National Council on Archives (NCA) format] or LC Name Authority File. Problems will therefore occur if a user is searching across different databases.

There are two fundamentally different approaches to identifying authors[111]:

String matching. Although string matching can use a number of other sources of information too, like human error detection, it is fundamentally about detecting name variations in a large database of names.

(Locally) controlled assignment of identifiers. In this procedure the authors of a new publication receive a unique ID only after humans have identified the author. Assignment of numbers is (preferably) locally controlled, e.g. by librarians, because they know, or can find, the authors. The assignment of IDs requires a central server that stores author identification numbers, together with information that makes identification possible. The server would send the IDs back if it is a known author, or issue a new id if the author is new.

11.8.1 Repositories

Name authority is a particularly significant problem for repositories. It appears that many people depositing materials in an institutional repository will not be represented in library name authority files, because they have not produced books or other materials which have been catalogued[112].

In addition, repositories rely on a large proportion of self-archiving, where authors are tasked with inputting their own information and barriers to data input must be kept low. It is unlikely that many repositories offer a facility for 'picking' names from an authoritative source. There are also issues of multiple authorship, often with authors at different institutions.

The issue of name authority for repositories was discussed at the JISC Information Environment and Digital Repositories Workshop in May 2006[113] and recognised as a significant problem. Eprints UK intended to offer some name authority control, but the right sources were not available. It was felt that the area requires investment and experimentation into metadata sets and authority control. There was a suggestion that HESA identifiers might be the building block for the UK, perhaps through a National HESA registry.

11.8.2 OCLC Research LC Name Authority Service

The OCLC name lookup web service[114] grew from plans to have a service that could be used to verify names for institutional repositories. It uses a matching algorithm for name lookup. The LC control number is used as the identifier. The display is identical to the LC MARC display.

11.8.3 SURF DAREnet 'National Author Thesaurus'

DAREnet is the network of Digital Academic Repositories. It is coordinated by the SURF Foundation, and includes all Dutch universities and several other academic and research organisations.

A DARE project to construct a 'National Author Thesaurus' is currently underway. Development is being carried out by OCLC Pica. A testbed version is being developed at the Groningen University, to be completed by June 2006. Although termed a thesaurus, in practice it is in fact a name authority list.

The National Author Thesaurus is based on a version of the thesaurus which formed part of the Dutch union catalogue. Because individual articles were not catalogued (only journals) this thesaurus only covered about 50% of Dutch authors. In order to populate the remainder of the thesaurus, author files in Metis (the institutional research registration system) are being matched against the Pica thesaurus. Coverage will therefore increase to 90% using automated processes. The remaining 10% will be input manually.

Numbers are assigned to unique individuals. To find the correct author, the thesaurus contains extra information about the author, such as date of birth, department he or she works for, etc. The immediate goal of this project is to numerically identify the authors from Groningen uniquely. The planned follow up is to extend it to all universities in the Netherlands. The next phase would be to link this work to already existing Name Authority lists[115].

The roll out of the new system is planned for Autumn 2006 and will be completed before the end of the year. The project is also looking for similar international activities for potential interworking.

11.8.4 The National Archives: National Name Authority files

The National Archives (TNA) was formed in April 2003 by merging the Public Record Office (PRO) and the Historical Manuscripts Commission (HMC). The former HMC developed the indexes to the National Register of Archives (NRA) to form the basis of National Name Authority files for persons, businesses and organisations. The current version of the NRA has fields for all the data elements required by the International Standard for Archival Authority Records (Corporate, Personal and Family) ISAAR(CPF)[116] standard, and some work has been done on modelling the content, look and feel of National Name Authority files based on this data structure.

The key issue with name authority files is generating the initial data to populate them. Archivists have always recorded more detail than libraries in name authority files, finding this necessary in order to distinguish between names. The NRA has some 180,000 standardised corporate, personal and family names, each of which needs to be developed from the current skeleton record into a full record by the addition of content and links. There are potentially many thousands more (including some on A2A (Access to Archives)[117]). Developing the name entries in the index into full authority records is a labour-intensive process, and has so far proved an insuperable barrier to the NRA indexes being launched formally as name authority files. Funding is unlikely to be available within TNA in the foreseeable future.

In order to progress development, TNA is keen to collaborate with JISC and other interested bodies. TNA is willing to provide leadership and technical expertise to support the initiative. Preliminary discussions have already taken place with a range of organisations including JISC, The Arts and Humanities Research Council, The Arts and Humanities Data Service, MLA, The Heritage Lottery Fund. It is also planned to include The British Library in discussions.

Another low-cost option might be to develop a moderated Wiki approach to the preparation of the records. This would enable remote (and international) individuals to submit information for inclusion in the records

It is interesting that name authority files were originally felt to be useful for archive cataloguers and for user searching. However they are now seen as valuable resources in their own right. The Oxford Dictionary of National Biography [118]is an example of this.

11.8.5 Other initiatives

Elsevier's Scopus Author Identifier is an initiative aiming to automatically match variations of an author's name and distinguish between authors with similar names[119].

At an international standards development level, the IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR) aims to study the feasibility of an International Standard Authority Data Number[120].

An International Standard Party Identifier (ISPI)[121] for people and organisations has been proposed and an outline developed. The next step is for NISO to take this on formally as a Work Task.

11.8.6 JISC Future Strategy

JISC does not currently have any specific activity in the area of name authority; it is timely to consider initiating some focused work.

- ***Recommendation: name authority for repositories requires further investigation, including the option of HESA identifiers.***
- ***Recommendation: the option of developing a UK name authority should be investigated.***
- ***Recommendation: JISC should work with other interested bodies including the British Library, and consider harnessing the enthusiasm of The National Archives (TNA) to lead a collaborative UK name authority effort.***
- ***Recommendation: collaboration with the SURF DAREnet name authority initiative should be explored.***

12 Shared Infrastructure services in context

The shared infrastructure services component of the JISC IE cannot be viewed in isolation but needs to be considered in relation to the other components of the IE and to other sectors within the UK and beyond. This integrated vision of the future is articulated by Atkinson et al[122] as:

> *"In the future a pervasive digital infrastructure will allow computing facilities to be always available via a heterogeneous range of devices. The infrastructure will seamlessly combine reliable high-performance computing and communications networks and variable low-performance embedded or portable devices with integrated wireless facilities. This will connect scientists in resource-rich labs to field scientists with limited resources or to remote automated experiments to for a distributed ubiquitous system. The supporting infrastructure will need to be open to all legitimate users, promote heterogeneity and be extremely flexible. Resources will vary in their availability, their certification of quality and their reliability."*

Shared infrastructure services are integral to the successful delivery of the JISC e-infrastructure, which a recent JISC briefing paper[123] describes as comprising the technology and organisations that support research carried out through distributed regional, national and international collaborations that utilise large data collections, advanced ICT tools for data analysis, large scale computing resources and high-performance visualisation. It goes on to state that it is the **integration** of all these elements – networks, grids, data centres and collaborative environments together with elements such as supporting operations centres, services registries, single sign-on, certificate authorities, training and help desk services – that defines the e-infrastructure.

However, shared infrastructure services also need to be seen in the context of the wider information landscape outside JISC borders.

Research Councils UK (RCUK)[124] is a strategic partnership through which the UK's eight Research Councils work together to champion the research, training and innovation they

support. As the main public investors in fundamental research in the UK, RCUK works alongside the Office of Science and Innovation (OSI)[125] to support the UK academic research and to ensure the best investment of public money in research. RCUK's recent statement on access to research outputs[126] reaffirms its belief in the value of repositories as a means of improving access to the results of publicly-funded research and encourages UK researchers to deposit their outputs in e-print repositories. JISC is not only supporting this through its Digital Repositories Programme and the Repositories and Preservation strand of its capital programme, and its support of the development of the UK PubMed Central repository, but also through its investment in shared infrastructure services.

JISC is also working with the Office of Science and Technology (OST), which supports the Government in developing and implementing its domestic and foreign policies for science and innovation. OST work includes investing with the Research Councils in research, research infrastructure and knowledge transfer and promoting international partnerships in research, science and technology. OST recently set up several working groups to look at e-Infrastructure; JISC co-funded the original roadmap report with OST and led some of the working groups and is currently writing up the final summary for OST. The working groups were variously led by the JISC, the Research Information Network (RIN)[127], and the BL. A set of draft reports has been produced on the following areas:

1. AAA, middleware and DRM
2. Networks, compute power and storage
3. Preservation and curation
4. Search and navigation
5. Data and information creation
6. Virtual research communities

These draft reports have not yet been released but JISC is currently drafting the synthesis report which will bring together all of the individual reports. It is worth noting that some of the points made in these reports could be said to depend on elements of the JISC shared infrastructure services being in place.

Beyond this there are further areas of the information landscape within the UK – the NHS, e-government, schools and cultural and heritage institutions – and their counterparts in other countries world-wide.

The NHS[128] is very supportive of the idea of shared infrastructure services in general and, more specifically, it would see potential benefits in the use particularly of JISC IE resolver services, identity management and document delivery. Work is currently going on to establish a National Knowledge Service (NKS), which would have shared services at its core. NHS Single Sign On has similarities with Shibboleth and it is expected that NKS would be procuring elements of shared services over the next couple of years. The NHS would be interested in collaborating with JISC in developing tools and services, and in being testbeds or pilot implementers of such tools and services.

The Strategic e-Content Alliance (SEA)[129] is a three-year initiative funded as part of JISC's Capital Programmes, running from March 2006 to March 2009. JISC is taking forward this work in collaboration with a number of key public sector organisations: The British Library[130], the BBC[131], British Education Communications and Technologies Agency (BECTa)[132], the UK e-Science Core Programme[133], the Museums, Libraries and Archives Council (MLA)[134] and the NHS National Library for Health (NHL)[135].

It aims to build on the first phase of work – the Common Information Environment (CIE)[136] – to build a common information environment where users of publicly funded e-content can gain best value from the investment by reducing the barriers that currently inhibit access, use and re-use of e-content. The vision is to achieve this through providing a set of principles and guidelines for best practice that will enable key public sector organisations to collaborate and co-ordinate their e-content activities to make best use of the limited funds available.

It will be important that there is interoperability between any e-infrastructure components that support Grid and e-Research and the JISC IE. The IE approach could usefully include these

aspects, once specific interoperability approaches have been agreed. For example, it might be agreed that if there is a science service registry, then it has to work with the IESR. Similar agreements might be possible for e.g. preservation services.

In addition to its participation in SEA, the MLA is itself is funding development of shareable services under its Digital Initiatives and Knowledge Web strategy and there is a perceived need for more piloting of service prototypes in all types of institution, including those outside the HE sector.

The BL also shares interests with JISC, values information on JISC initiatives and is positive about collaboration with JISC when appropriate opportunities arise; the appointment of a JISC/British Library Partnership Manager has been valuable in identifying such opportunities. At the workshop held as part of this study, it was noted that the size and range of BL activities means that there are many areas of overlapping interest but the following areas were identified as having the most potential for collaboration at present.

- Institutional repositories – the BL is currently working on a JISC-funded project to provide a repository of e-Theses and has its own SHERPA directory. All underpinning standards would be of interest.
- Digital policy management – the BL has also worked with Rightscom on similar issues to JISC and is interested in ONIX for Licensing terms and there are potential opportunities for piloting collaborations. The BL is already involved in OLT initiatives via the JISC PALS Metadata and Interoperability Working Group[137].
- Identifiers – the BL is aware of the critical importance of unique and persistent identification in the context of digital resources but has not identified any specific collaborative activity.
- Metadata registries – this is of interest to the BL as a tool for managing multiple metadata formats internally. This interest extends to the BL's involvement with The European Library, which also has a metadata registry, and the European Digital Library.
- Name authorities – the BL may be interested in any initiatives in this area.
- Versions – this is of interest to the BL.

The brief review above of the wider information landscape shows that there is already collaboration and co-operation between the JISC and other organisations. In order to gain full benefit from the willingness to co-operate, it is important that the JISC maintain regular contact and consultation with a variety of other organisations in order to identify areas of potential collaboration and synergy before multiple individual projects are set up for the same aim.

- ***Recommendation: JISC and shared infrastructure services need to develop a common understanding of the project lifecycle, so that stakeholders can be assured of continuity within the limitations of short term funding cycles imposed on government agencies such as JISC.***
- ***Recommendation: JISC need to promote a better understanding of the IE in the wider community, so that larger vendors are more prepared to work with JISC services.***
- ***Recommendation: JISC should maintain regular contact and consultation with other key organisations, both in the UK and internationally, in order to facilitate early identification of potential collaboration and synergy.***
- ***Recommendation: JISC should work with other key organisations to ensure interoperability between any e-infrastructure components that support Grid and e-Research and the IE. The IE itself should be developed with such interoperability in mind.***

13 Key Issues and community concerns

Having considered what shared infrastructure services components of the JISC IE have been defined in the IE architecture and reviewed the current status of projects and services, both

within JISC and outside it, the question arises ‘where do we go from here’. During the review, as a result of background literature, interviews and workshop discussions, some key issues have emerged.

13.1 External contacts

It is important for the JISC IE not to be developed in isolation. External contacts are important and cross-sector working should be encouraged. The JISC IE and the SEA have the same aim of connecting user and required resources in a seamless fashion. The British Library, MLA, The National Archives, BIC and the publishing trade, the health service and Becta have all indicated an interest in working with JISC and these initial indications need to be followed up in a systematic way to achieve worthwhile collaborations.

13.2 Developing shared infrastructure services

The development community have raised a number of issues on this point. Discussions with JISC indicate that these issues are sometimes, but not always, the result of a lack of information or mistaken assumptions. A key point here is that JISC needs to work on communicating more effectively to the community.

One example of a communication issue is the regular references in documentation to ‘JISC services’. JISC should clarify if it is just referring to MIMAS and EDINA, or a wider group of services.

JISC funds some work on a research basis; the outcome might be a ‘failure’ but is useful in identifying an approach that does not work or a potential alternative approach that might work better. The community accepts that not all work is funded with an expectation of progression to service.

In the case of projects where prototypes have yielded positive results, JISC has funded further phases of work. However, in these cases, there is sometimes a difference in perception of status with JISC seeing the work as part of a (probable) progression to service, while the project perceives it as a case of continually finding new funding with consequent struggles to retain staff expertise and effort allocation over the gaps between phase funding.

Where a project has developed to the point that it is able to interest other parties in using its tool or service, difficulties tend to arise concerning project status. It is difficult to convince external organisations and services to invest their resources in a service whose long-term future is uncertain.

Even with external buy-in to specific services, it may be that some core funding is required to support some services, especially in the early stages of full service provision.

Although in an ideal world, JISC would simply be able to commit to fund every project that delivered real benefits to the user community as a service, the reality is otherwise. Since government agencies such as JISC are themselves on a short-term funding cycle, it is difficult to commit too far ahead. JISC should be able to address this more effectively as their development to service and project lifecycle work develops.

For some projects, there is an issue over data input. Initial project work may be funded to populate a database sufficiently to demonstrate principles and that an approach will work. However, population of the database with further data may be required in order to take the tool or service forward; funding for this is often not available to the project or to potential collaborators who have appropriate data, thus stalling progress.

Designating a project for transition to service brings its own set of issues. Projects will need to consider, amongst other things, how the new status affects personnel, software design, scalability, and tie-ins with other services and ‘customers’. It might be useful for JISC to consider funding a small study to look at this aspect of service development.

JISC is now addressing some of these concerns with its work on developing a project maturity scale. The scale has five points:

- Exploratory project

- Project with identified service potential
- Ready for consideration for transition to service
- Approved for transition to service
- In full service

Definitions and guidelines for application of the scale points are still being drafted.

It is also recognised that JISC is acknowledging project concerns by its planned changes to methods of awarding continued funding.

13.3 Business models and cost-benefit analysis

As services move towards full service provision, there is a need to develop business models. However, there has been little guidance from JISC as to the methods or metrics that should be used in this work. The academic sector has limited experience of running business-based services and full commercial viability may not be achievable for all services. JISC should think through what commitment it can make to provide a level of baseline support for any shared service.

Cost-benefits will include those that are measurable (e.g. increased use of a service) and those that are intangible, such as the trust in a resource that is built up through reliable and quality delivery of service.

JISC and the HE sector already operate in a mixed economy environment, making use of some proprietary products and services. It may be that some shared infrastructure services developments can be exploited on a commercial basis, and semi-commercial models whereby (e.g. the NHS) buy services, thus effectively sharing the cost with HE/FE. JISC should keep all options open regarding service usage. It should be noted that the business model becomes more critical when looking at service usage outside the JISC funding scope, as would be the case with international and cross-sectoral collaboration.

There is also a need to specify the context in which cost-benefits are calculated. It is obvious that cost-benefits to the academic sector need to be considered, but how far outside this should services be looking and which areas should have priority? Within the UK, services could be of use to schools, the health service, public sector cultural and heritage organisations, local administration and national government services, and the CIE. Should services also look at cost-benefits to sectors and organisations outside the UK, and should the commercial sector also be considered?

Once a shared service is in operation, there will be a need to monitor its usage in some way. For example, those institutions who have installed OpenURL Resolver have noticed an increase in accesses requiring this service. Monitoring should be in an appropriate form for the service monitored and should ideally be automated and not rely on large amounts of human effort.

13.4 Lack of community input to shared infrastructure services programme

There is a feeling that the JISC shared infrastructure services programme has been designed from the top down and with little bottom-up response or input, although sometimes such input has initiated developments. The OpenURL Router is an example of a service developed in response to need from institutions and not identified through the top down approach.

Security and sustainability are clearly of importance to stakeholders. There is concern that there is no clear way for the community to articulate concerns, which include the transparency of funding decisions, and scalability issues.

Although JISC itself feels that there is input through its various programme committees, whose membership includes people working ‘at the coal-face’, it has also decided to fund an ‘IE Institutional Landscaping Survey (Invitation to Tender issued on 3rd July 2006) which aims to establish how ready institutions are to take up services, what services they want, and which are ready and willing to be early adopters; it will also look at barriers to adoption, such as the impact of legacy systems.

13.5 Promoting shared infrastructure services

Even where institutions want to use shared infrastructure services, they may have a lack of understanding of what these services can do and/or be unclear about how they can do so. The JISC should not only recognise that shared infrastructure services are a similar 'common good' as that provided by the network (albeit not as common or as shared as the network).

There is a strong feeling among the projects consulted that shared infrastructure services are the 'poor relation' within the whole of the JISC work programme and there is a need for advocacy in this area. By their nature M2M services are invisible and potential users need to be convinced of the value of these 'invisible' services and JISC needs to demonstrate the dangers of not taking them on. JISC itself is investing funding in this area and has undertaken this report to help address such problems.

JISC needs to make it explicitly clear how shared infrastructure services can and should be used so that it is easier to sell these within an institution. It is felt that shared infrastructure services should be embedded in other programmes; this could be, for example, through requiring that new repositories funded under JISC calls should make use of any appropriate available tool or service, at the least in a pilot capacity. The new JISC IE testbed may help in this.

13.6 Delivery at institutional level

It will be important to distinguish between institutional and non-institutional shared infrastructure. This review has concentrated on a shared infrastructure that is delivered outside the institution and developed at network level, as JISC seeks to identify what it can deliver at a level that institutions cannot manage individually, but it is also useful to consider how the functionality of those services is best delivered within institutions. For example, developing a service registry at network level is a good way to maximise results from limited expertise and funding, but to be truly useful the functionality must be delivered within an institution; this could be through web service look-up or by harvesting the database and serving it locally. However, it is not so clear how other shared services will be delivered within an institution. In some instances where shared infrastructure services have been developed at network level, it might be most appropriate to deliver through individual implementation at institutional level rather than through a network service. This would be useful where localisation of functionality is valuable; OpenURL knowledge-bases are an example of a shared service that is served locally to allow for localisation.

13.7 Digital Policy Management

As previously stated in section 6, the development of digital policy management strategies and services cannot be carried out in isolation; it must be tied in to work on authorisation and authentication and it will be advantageous to also work collaboratively with content suppliers and the Access Management Federation[138] (due for launch in September 2006, this will provide next generation access management facilities to users and institutions across the UK, using Shibboleth technology).

Additionally, there may be issues related to licensing of content; if a licence is negotiated for specific usage, this may prevent its usage in one or more shared infrastructure service until a new licence is negotiated – which will add to the cost.

A related concern is continued access to resources when a subscription has been cancelled. Such access could be provided by continued access to ring-fenced sections of a resource (anything deposited before the subscription cancellation date), relevant content deposited with institution, or content deposited in a UK archive.

13.8 Provision of a testbed environment

In some cases it would be useful to trial a specific tool or service within a testbed environment. It has been suggested that a 'working IE in miniature' should be built that would enable testing

of the usefulness or otherwise of shared infrastructure services to real end-users and the services they use. More specifically, it was suggested that this should be a rich environment, built in a phased way, that recognises that there may be more than one terminology service, collections and services database, etc. Evaluation should be a major aspect of such an environment, looking at cost-benefits, alternative ways of providing shared infrastructure services, gaps, inter-service interoperability, and interaction problems thrown up by real M2M services working in a real environment on real service-end and user problems.

JISC has already decided to fund this work and an Invitation to Tender is about to be issued.

A testbed environment is not the only way to assess a tool or service and it will still be necessary to pilot tool and service prototypes in institutions. It is important that such piloting takes place in all types of institution, and not just in resource-rich institutions where internal technical support is to hand.

13.9 Software Quality Assurance and Testing

A view has been expressed that a number of JISC projects include the development of software, but what is developed may not be capable of supporting an operational service or be able to be put in the public domain as Open Software contributions. It has been suggested that the reason for these JISC-funded software being prone to system failures or becoming rapidly obsolete before being used in real services or by other projects, is the lack of a suitable software quality assurance and testing management system (SQATM).

Currently the JISC requests projects to follow its Guidelines for Software Development[139]; however, these are generic guidelines and do not clearly establish metrics or the means for making sure that best practices such as software coding and annotation standards are consistently applied through the software production life cycle, or mechanisms for the detection of errors and bugs, etc.

A software quality assurance and testing management system does not need to be a complex package of software or a commercial 'black box' and most of the tools or components required are already available to developers or are embedded in development environments used by developers. It would therefore be possible to develop a set of clear and specific standards, procedures and testing toolkits for JISC projects. Since such a solution needs to help the developer and not become yet another obstacle, the package should be sufficiently flexible to be adapted and used by large, medium and small software projects; this would require a background study to investigate questions such as: whether all JISC software requires such a package, what are common problems in software development, and how could such a package be introduced into the existing JISC processes?

Other issues about software should not be ignored. Services built on software must be robust enough to function at times of heavy use (e.g. to the level of a million hits a day). Software code should not be lost when a project finishes; there should be guidelines on deposit (e.g. in SourceForge[140] and GForge[141]) and preservation. In addition, good quality software should also be supported by materials to support institutional adoption.

- ***Recommendation: JISC should fund a small background study prior to commissioning (or inviting to tender) a software quality assurance and testing management system (SQATM) package.***
- ***Recommendation: JISC should produce guidelines regarding the deposit and long-term preservation of software in appropriate locations.***
- ***Recommendation: Commissioning of any (pilot or production) should include, in addition to working software) a complete package of materials (e.g. the 'how to' guide for installation, glossy brochures for promotion, phone support, support/users e-list) to promote and facilitate institutional adoption.***

13.10 Supporting poorer institutions

It has been noted that the uptake of certain tools and services may be restricted by their cost. Institutions in the higher education sector vary in their capacity to fund buy-in, even where there

is interest and desire to use a particular tool or service. This lack of available funding is even more apparent in the further education sector.

JISC should therefore develop a strategy that would capitalise on the work done by projects to develop a tool or service and support its use within the whole of the HE and FE community. There are a number of ways in which this support could be provided.

(a) Develop generic versions of specific services, such as OpenURL Router, that could be made available to institutions that lack sufficient budget to buy the 'real' service.

(b) Fund the interface customisation that might be required for installation, as might be the case with GeoCrossWalk.

(c) Require (and fund) services to develop model generic clients and provide detailed installation guides that would help institutions with limited access to technical support.

13.11 Supporting standards

JISC rightly promotes the use of national and international standards as appropriate to support interoperability and flexibility but does not act in a standards making capacity. However, the JISC community are variously involved in creating and developing standards, with varying levels of support from institutions. It is not clear whether shared infrastructure services are funded for this, even when a new standard is required, as was the case for collection description. There is also the issue of UK input to the development of international standards. Some people contribute to this 'out of hours' because of an interest, while others have institutional permission to work on something as long as the institution incurs no financial expense.

The JISC Standards Framework, currently being developed by UKOLN, is described at the beginning of section 11.

JISC is also supporting Dublin Core Metadata Initiative (DCMI)[142] work as joint funders (with MLA) of the UK's role in the DCMI Affiliate Programme. UKOLN is acting as 'managing agent'[143].

14 The Way Forward

This study has looked in some detail at the current state of JISC shared infrastructure services and their relationship to the JISC IE architecture. It is clear that although work is going on to provide the shared infrastructure services that the architecture envisages, each area is at a different stage of development and offers varying opportunities to use tools and services developed outside the HE sector and/or to collaborate with other organisations to develop tools and services.

The previously mentioned *Digital Repositories Roadmap* by Heery and Powell looks at the potential landscape for repositories in 2010, and a similar approach is adopted here.

14.1 The Vision

In 2010 the content of the JISC IE will be both bigger and richer. There will be more repositories, serving a variety of needs and users and offering a richer scholarly communication environment based on open access to, and re-use of, scholarly materials. Not only will these be available but it also it will be possible for supporting metadata to be exposed to applications and for services to be based on materials held in repositories. Repositories are expected to become more embedded in the information landscape, with better interoperability with systems used to support learning and teaching, as well as authoring tools, other repositories, portals and library systems.

Large scale scientific collaborations will be supported by grid architecture and the infrastructure and services that can be built on it. Grid services and applications will need to use shared infrastructure services and are likely to place a heavy demand on the services used, with the requirement that such services are robust and can sustain heavy usage.

The OST e-infrastructure reports envisage that there will be more digitised content that, together with deployment and exposure of rich metadata, will enable wider access to the resources of archives, libraries and museums. More effective use of metadata will improve abstracting and indexing services, provide better access to newer forms of resource such as video and enhanced provision of full text content (free at point of use). Individuals workflow patterns will be made more effective through the use of linking between articles and author emails and home pages, citation lineages for articles, influence lineages for topics, interconnection between commentary and blog space with research space, and the personalisation that will be built into systems. Search and navigation tools will assist researchers working on and across the boundaries between traditional disciplines and subject areas, and services will be widely available to facilitate searching across languages and to enable search and navigation by space, time, person and concept. Data will be captured systematically as part of researchers' workflow processes; the use of metadata standards will facilitate the effective functioning of M2M services. M2M services will also mitigate the barrier effects of IPR issues, through the use of machine-readable licence registries.

14.2 How can JISC move on?

Although this study looks at what the JISC needs to do, it is not the only player. Some objectives will be better achieved by using tools and services developed by other or through collaboration with partners in different sectors. The JISC often has a role as the seed-corn starter, with other players moving initiatives on. Another role for the JISC in this 'mixed economy' model is to influence other providers (e.g. OCLC) through collaborations and other forms of input into development processes.

14.3 Working with shared infrastructure services and collaborations

There is a need for JISC to work on its strategy regarding shared infrastructure services development. This includes developing a process lifecycle for tool and service development that is tied in with the maturity scale currently in draft form; this would, of course, not guarantee that every exploratory project would progress to full service. The lifecycle and maturity scale needs to be promoted in the community, such that existing projects and potential bidders for projects are clear about the JISC process and the implications for their project.

Although the JISC funding of shared infrastructure services is primarily for 'seed-corn' development, there is a need for acknowledgement that some services may need interim funding as they move to service delivery under a business model, and even that some services, due to their nature, might require some element of core funding in the longer term. Projects also need more support in developing business case models, in particular with guidance on the metrics to be applied and a way of valuing intangible benefits.

14.4 Technical issues

JISC needs to review its strategy and guidelines regarding software. Generic guidelines are in themselves insufficient to guarantee good quality software and JISC should look at further work to provide a toolkit of standards and tools, and the development of guidelines on the provision of supporting documentation and on the long-term preservation of software; these should be compulsory for projects.

Work in the area of digital policy management cannot be divorced from authentication and authorisation issues, although these are outside the remit of this study. Collaboration with content providers and BIC in initiatives such as ONIX for Licensing Terms is vital to enable the fullest access to resources, whilst working with the implications of IPR.

The planned testbed environment will provide shared infrastructure services with an opportunity to trial prototypes outside the institution or environment in which they were developed, to investigate how they would function in the real service delivery mode. However, piloting services in other environments in collaboration with partners in other sectors will still be a valuable exercise in moving projects to full service.

14.5 Advocacy

Now that some shared infrastructure services are at or nearing service delivery level, it is important that the JISC develops a strategy to promote shared services both within the HE/FE community and outside.

The scenarios developed for this study could be used as the basis for a range of publicity materials to promote shared infrastructure services in general or for a specific tool or service.

The JISC IE will function in the context of the wider information environment. Collaboration is an effective way of leveraging effort from a variety of sources to benefit all. The JISC should actively seek more collaboration with other sectors and organisations, several of which are already positive about such partnerships. Collaborations can be at any level of project maturity, but would be of great value at piloting and trial service stages.

15 Appendix 1: JISC Development and Service Maturity Scale

Draft version 0.3 dated 5th June 2006.

The proposed maturity scale is related to the types of projects funded through JISC Development Programmes and to the committee governance structure and processes through which decisions are reached. It is intended that this maturity scale should in the first instance be applied to the JISC Information Environment Inventory, but potentially it could have wider application, particularly in development to service transition processes. The scale has five points:

- Exploratory project
- Project with identified service potential
- Ready for consideration for transition to service
- Approved for transition to service
- In full service

In practice, there may be some difficulty in saying when a particular area of development has "identified service potential" – there are draft guidelines currently being piloted for decision points between the other points on the scale.

15.1 In full service

These are services that:

- have software that has been made robust and packaged for ease of maintenance;
- have defined benefits, an approved business case and an assured funding stream (subject to normal cycles of review) to cover management: making them available (hosting, etc.), support (e.g., help desk, training, small-scale fixing), and promotion;
- if JISC Services, are subject to the terms of a service level agreement, and/or other agreed management procedures.

Further development is almost certain to be required in all cases. Depending on the scale of then development required, this may be funded from an ongoing budget-line for development (usually covering only small-scale developments, but could be larger technology refresh where this can be anticipated) or through new projects within Development Programmes (usually larger-scale and innovative developments).

15.2 Approved for transition to service

The outputs from a project or a group of projects have been approved for transition from development project to service status, based on review by the appropriate JISC sub-committee. Two stages (or tranches) of development follow the initial approval. During the first phase the outputs will be brought up to the standards implied in 1, above. This process of making robust

and packaging software and related infrastructure is sometimes termed "productising", so this is a "productise projects tranche". In addition, a refined business case will be produced, perhaps amalgamated from the business cases for individual projects. Refinement of benefits profiles may also be required. The second tranche involves quality assurance and evaluation of the productised service. In practice, this might be one project with a review point part way through.

15.3 Ready for consideration for transition to service

Sufficient development has been done for the outputs of a project, or group of projects to be considered for transition from project to service status. These are outputs from development projects that have been identified as contributing to the planned programme outcome: "develop services, infrastructure or applications that may be used at departmental, institutional regional or national level". Consideration will be given via review by the appropriate JISC sub-committee, involving examination of business cases and benefit profiles. Projects at this level are sometimes called pilot services; the intension of such projects is to build services or tools. Services may be software-based, or may be advisory services.

15.4 Project with identified service potential

Some development has been done unless a project is at a very early stage of a first-phase. However, more development is required, either within the current project, or in a further project phase. Projects at this level are sometimes called trial services or software, prototypes or demonstrators.

15.5 Exploratory project

Exploratory projects are at an early stage in the exploration of new technologies and approaches. Proof-of-concept demonstrators may be built. Although they may produce "best practice" guidance, they not expected to produce outputs that would lead to services, or implemented software. However, their results are likely to inform future funding decisions.

16 Appendix 2: List of people consulted during the study

16.1 By Ann Chapman and Rosemary Russell

Julie Allinson (UKOLN)

Ann Apps (MIMAS)

Adrian Brown (The National Archives)

Rachel Bruce (JISC)

Peter Burnhill (EDINA)

David Giaretta (DCC)

Brian Green (BIC & ONIX)

Rachel Heery (UKOLN)

Bill Hubbard (SHERPA-ROMEO)

Nick Kingsley (The National Archives)

Traugott Koch (UKOLN)

Lesley Mackenzie-Robb (Becta Vocabulary Studio)

Dennis Nicholson (HILT)

Christine Rees (EDINA)

James Reid (EDINA & GeoCrossWalk)

Chris Rusbridge (DCC)

Emma Tonkin (UKOLN)

Leo Waaijers (SURF DAREnet) (by email)

Scott Wilson (CETIS)

16.2 By Mark Bide

Naomi Korn (Naomi Korn Copyright Consultancy)

Charles Openheim (Department of Information Science, Loughborough University)

16.3 Workshop participants [29th June 2006 at The King's Fund, London]

Ann Apps (MIMAS)

Chris Awre (University of Hull)

Neil Beagrie (BL/JISC)

Rachel Heery (UKOLN)

Amanda Hill (MIMAS)

Dennis Nicholson (HILT)

John Paschoud (LSE)

Andy Powell (Eduserv)

Phil Purdy (MLA)

Stephen Rankin (DCC)

James Reid (EDINA & GeoCrossWalk)

17 Appendix 3: Key Issues identified at workshop

- External contacts are important – e.g. the Strategic e-Content Alliance
- Need for consultation with other sectors – e.g. BL. Clear consultation process.
- Cost benefits (no guidance from JISC on methods or metrics to assess a service or tool)
- How far should other sectors / countries be included?
    - Benefits for other services – e.g. schools, Strategic e-Content Alliance, commercial (prioritise)
- Everything seems to be top down – little bottom up response; e.g. OpenURL Router driven by a need from institutions.
- More piloting of service prototypes needed in all types of institution (including non-HEI)
- What's missing – mini-working environment/testbed (stimulates people to think) [not everyone agreed]
- Need to tie in DPM to authorisation – can't be separated… (Mark Bide does comment on this – highlight this)
- Implication of transition from project to service – how it affects personnel, software design – get tied into things… This could continue in future 'Research' Call.
- Projects intended to be prototype services v. projects intended to contribute to learning (JISC expect some of the latter to fail)
- It would be useful to explore the hypothesis that there is no need for [JISC?] shared services

- Security, sustainability, stakeholders (top down – no clear way for community to articulate concerns – transparency of funding decisions), scalability
- Even where institutions want to use, not clear how – lack of understanding of what SS will do – they need to be understood by perceived stakeholders
    - don't necessarily have resources to enable this – need support/funding
    - SS should include means to use them eg model generic client, installation guide (but resources needed)
- Interoperability – flexibility – how SS can adapt in future to new tech – standards
    - who funds internat standards devel?
    - not explicit whether services already funded for this – devel takes years
- Promote & explain M2M services – people don't understand invisible stuff
- Status of SS – poor relation to other JISC services – should be embedded in other services – advocacy; embedding needs to be built into SS from beginning
- Licencing for content issues – market is via SS – negotiated for that purpose, could be later probs – extra cost if need to renegotiate
- Copyright
- Educating content suppliers
- Assessing intangible benefits
- Explore whether should be exploited on commercial basis – keeping options open…
- Other 'commercial' models eg NHS may buy services – means shared costs with HE/FE

18 Author contact details

Ann Chapman
UKOLN
University of Bath, Bath BA2 7AY

Tel: 01225 386 121
Email: a.d.chapman@ukoln.ac.uk

Rosemary Russell
UKOLN
University of Bath, Bath BA2 7AY

Tel: 01483 560 342
Email: r.russell@ukoln.ac.uk

19 References

[1] Carpenter, Leona. Moving forward with shared services. Annex A. [PDF supplied to authors]
[2] Investing in the Future: Developing an Online Information Environment. URL: http://www.jisc.ac.uk/index.cfm?name=ie_home
[3] The JISC Information Environment Technical Standards. URL: http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/standards/
[4] Powell, Andy. Mapping the JISC IE service landscape. Ariadne Issue 36, July 2003. URL: http://www.ariadne.ac.uk/issue36/powell/
[5] OpenURL Router. URL: http://openurl.ac.uk/doc/
[6] The Shibboleth Project. URL: http://shibboleth.internet2.edu/

[7] The Handle System. URL: http://www.handle.net/
[8] The Digital Object Identifier System. URL: http://www.doi.org/
[9] PURLs. URL: http://purl.org/
[10] The e-Framework for Education and Research. URL: http://www.e-framework.org/
[11] Intrallect Study. URL: http://www.intrallect.com/drm-study/DRMFinalReportv2.pdf
[12] JISC Rights in Digital Environments Workshops. URL: http://www.jisc.ac.uk/events_ipr.html
[13] Korn, N., Oppenheim, C. and Picciotto, S. *Legal and policy issues session*, at the 2nd Digital Repositories Programme meeting (march 2006).
URL: http://www.ukoln.ac.uk/repositories/digirep/index/legal and policy issues cluster session 2006-03-7
[14] TrustDR. URL: http://www.uhi.ac.uk/lis/projects/trustdr/
[15] Rights and rewards. URL: http://rightsandrewards.lboro.ac.uk/
[16] Project COUNTER. URL: http://www.projectcounter.org/
[17] SUSHI. URL: http://www.niso.org/committees/SUSHI/SUSHI_comm.html
[18] GeoCrossWalk – Example Use Cases. May 2006 draft; finished report not made available at time of writing
[19] HILT Phase III Use Cases, in M2M Pilot Demonstrator Project Proposal. URL: http://hilt.cdlr.strath.ac.uk/hilt3web/AboutHilt/HILTM2MPilotbid.pdf
[20] IEMSR Phase 2 User Requirement documents. URL: http://www.ukoln.ac.uk/projects/iemsr/2wp2/
[21] "Use the IESR". URL http://iesr.ac.uk/use/
[22] "A Middleware Registry for the Discovery of Collections and Services" by Ann Apps, presented at NCeSS2005. URL: http://iesr.ac.uk/pubs/
[23] Rights in Digital Environment report. URL: http://www.jisc.ac.uk/index.cfm?name=events_ipr
[24] Scoping study into Digital Rights Management. Study and appendices. URL: http://www.jisc.ac.uk/index.cfm?name=prog_middss_studies
[25] Scoping study into Institutional Profiling and Terms and Conditions Services. URL: http://www.jisc.ac.uk/uploaded_documents/CMSS-Shaw1.pdf
[26] JISC IE Architecture Usage Scenarios. URL: http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/scenarios/)
[27] Powell, Andy. *The JISC Resource Discovery Landscape: a personal reflection on the JISC Information Environment and related activities*. May 2005. URL: http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/resource-discovery-review/
[28] Powell, Andy. *A 'service oriented' view of the JISC Information Environment.* November 2005. URL: http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/soa/
[29] Usage Scenarios for the IE Metadata Schema Registry. URL: http://www.ukoln.ac.uk/projects/iemsr/wp2/usage
[30] O'Looney, John. The Future of Public-Sector Internet Services: Pt. I. September 2001. URL: http://www.govtech.net/magazine/story.print.php?id=5806
[31] Bonett, Monica. Personalization of Web Services: Opportunities and Challenges. Ariadne 28, June 2001. URL: http://www.ariadne.ac.uk/issue28/personalization/
[32] Ferguson, Nicky, Schmoller, Seb, Smith, Neil. *Personalisation in presentation services: a report commissioned by JISC.* 4 August 2004. URL: http://www.therightplace.plus.com/jp/index.html
[33] Heery, Rachel and Anderson, Sheila. *Digital Repositories Review*. UKOLN, AHDS, 2005. URL: http://www.jisc.ac.uk/uploaded_documents/digital-repositories-review-2005.pdf
[34] Heery, Rachel and Powell, Andy. Digital Repositories Roadmap: looking forward. UKOLN, Eduserv, 2006. URL: http://www.jisc.ac.uk/uploaded_documents/rep-roadmap-v15.doc
[35] Swan, Alma and Awre, Chris. Linking UK Repositories: scoping study report. URL: http://www.jisc.ac.uk/uploaded_documents/Linking_UK_repositories_report.pdf
[36] Swan, Alma. Linking repositories scoping study. Presentation. URL: http://www.ukoln.ac.uk/repositories/digirep/index/Integrating_infrastructure_cluster_session_2006-03-27_report
[37] National s-Science Centre (NeSC) and e-Science Grid. URL: http://www.nesc.ac.uk/
[38] Report on 2nd Concertation Workshop on eInfrastructure, 14th December 2005, Bordeaux. URL: http://www.geant2.net/server/show/ConWebDoc.1631
[39] Permanent Access to the Records of Science. URL: tfpa.kb.nl/Proposal%20Research%20and%20Development.doc
[40] European Strategy Forum on Research Infrastructures (ESFRI). URL: http://cordis.europa.eu/esfri/
[41] U.S. Department of Energy. Experimental Program to Stimulate Competitive Research (DOE/EPSCoR). URL: http://www.er.doe.gov/EPSCoR/
[42] Internet Engineering Task Force (IETF). URL: http://www.ietf.org/
[43] NSF Cyberinfrastructure report. URL: http://www.adec.edu/nsf/nsfcyberinfrastructure.html
[44] JISC e-Infrastructure Programme. URL: http://www.jisc.ac.uk/index.cfm?name=programme_einfrastructure
[45] WWW 2006 Paper: A Contextual Framework For Standards. URL:

http://www.ukoln.ac.uk/web-focus/papers/e-gov-workshop-2006/html/
[46] Powell, Andy. Breakout briefing: persistent identifiers. The Digital Library and its Services, 6/7 March 2006, The British Library.
[47] OST Data Information and Creation Working Group: e-Infrastructure report 5: 20/20 Vision: an e-Infrastructure for the next decade [Draft report, not publicly available]
[48] The OpenURL Framework for Context-Sensitive Services. URL: http://www.niso.org/committees/committee_ax.html
[49] Apps, Ann and MacIntyre, Ross. Why OpenURL? D-Lib Magazine, May 2006. URL: http://dlib.ukoln.ac.uk/dlib/may06/apps/05apps.html
[50] Apps, Ann and MacIntyre, Ross. Why OpenURL? D-Lib Magazine, May 2006. URL: http://dlib.ukoln.ac.uk/dlib/may06/apps/05apps.html
[51] SFX. URL: http://www.exlibrisgroup.com/sfx.htm
[52] OCLC acquired the assets of Openly Informatics in January 2006
[53] The OpenURL Router. http://www.openurl.ac.uk/doc/
[54] ZBLSA project. URL: http://www.joinup.ac.uk/zblsa/
[55] OCLC OpenURL Resolver Registry. URL: http://www.oclc.org/productworks/urlresolver.htm
[56] Andy Powell email discussion, 23 June 2006.
[57] OpenURL CoinS: a convention to embed bibliographic metadata in HTML. URL: http://ocoins.info/
[58] Ball, Alex. *Briefing Paper: File Format and XML Schema Registries*. UKOLN, May 2006. URL: http://www.ukoln.ac.uk/projects/grand-challenge/papers/registryBriefing.pdf
[59] OAIS Reference Model. ISO 14721:2003
[60] Wotsit's Format. URL http://www.wotsit.org/
[61] File Format Encyclopedia. URL: http://pipin.tmd.ns.ac.yu/extra/fileformat/
[62] FILExt. URL: http://filext.com/
[63] National Digital Inforamtion Infrastructure and Preservation Program (NDIIPP). URL: http://www.digitalpreservation.gov/
[64] Sustainability of Digital Formats. URL: http://digitalpreservation.gov/formats/
[65] DCC RI RegRep: URL: http://dev.dcc.rl.ac.uk/twiki/bin/view/Main/DCCRegRepV04/
[66] CASPAR. URL: http://cordis.europa.eu/ist/digicult/caspar.htm
[67] The DCC Approach to Digital Curation. URL: http://dev.dcc.ac.uk/twiki/bin/view/Main/DCCApproachToCuration/
[68] PRONOM. URL: http://www.nationalarchives.gov.uk/pronom/
[69] Digital Record Object Identifier (DROID). URL: http://droid.sourceforge.net/wiki/index.php/Introduction
[70] Global Digital Format Registry. URL: http://hul.harvard.edu/gdfr/
[71] VERSIONS Project. URL: http;//www.lse.ac.uk/versions/
[72] EAST. URL: http://nssdc.gsfc.nasa.gov/nssdc_news/mar02/EAST.html
[73] Typed Object Model (TOM). URL: http://tom.library.upenn.edu/
[74] JHOVE (JSTORE/Harvard Object Validation Environment). URL: http://hul.harvard.edu/jhove/
[75] Australian Partnership of Sustainable Repositories (APSR). URL: http://sts.anu/edu.au/apsr/
[76] RoMEO. URL: http://www.lboro.ac.uk/departments/ls/disresearch/romeo/
[77] Autralian Research Repositories Online to the World (ARROW): URL: http://arrow.edu.au
[78] ONIX for Licencing Terms. URL: http://www.editeur.org/
[79] Book Industry Ccommunication (BIC). URL: http://www.bic.org.uk/
[80] EDItEUR. URL: http://www.editeur.org/
[81] Digital Library Federarion (DFL). URL: http://www.diglib.org/
[82] DFL Electronic Resource Management Initiative. URL: http://www.diglib.org/standards/dlf-erm02.htm
[83] EDItEUR and ONIX for Licensing Terms. URL: http://www.editeur.org/
[84] Association of Learned and Professional Society Publishers (ALPSP). URL: http://www.alpsp.org/default.htm
[85] JISC PALS Metadata Interoperability Projects. URL: http://www.jisc.ac.uk/index.cfm?name=programme_pals2
[86] IESR. URL: http://www.iesr.ac.uk/
[87] Amanda Hill. The Information Environment Service Registry: promoting the use of electronic resources. Ariadne Issue 40, 30-July-2004. URL: http://www.ariadne.ac.uk/issue40/hill/
[88] The Ockham Initiative. URL: http://ockham.org/
[89] Grimoires Registry. URL: http://twiki.grimoires.org/bin/view/Grimoires/
[90] OpenDOAR Directory of Open Access Repositories. URL: http://www.opendoar.org/
[91] Heaney, Michael. *An Analytical Model of collections and their Catalogues.* URL: http://www.ukoln.ac.uk/metadata/rslp/
[92] Shreeves, S.L. and Cole, T.W. *Developing a collection registry for IMLS NLG digital collections*. URL: http://purl.oclc.org/dc2003/03shreeves.pdf

[93] NISO Metasearch Initiative Collection Description Specification. URL: http://www.niso.org/committees/MS_initiative.html
[94] Baker, T. *et al. Principles of Metadata Registries.* White Paper of DELOS Registries Working Group. URL: http://delos-noe.iei.pi.cnr.it/activities/standardizationforum/Registries.pdf
[95] Powell, Andy. *JISC IE Architecture: Shared Services Development Plan.* Draft, 9 April 2003. URL: http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/ssplan/
[96] IEMSR. URL: http://www.ukoln.ac.uk/projects/iemsr/
[97] The European Library. URL: http://www.theeuropeanlibrary.org/
[98] The National Science Digital Library (NSDL) Metadata Registry. URL: http://eg2.ischool.washington.edu/registry/
[99] Dataset Acquisition, Accessibility and Annotation e-Research Technology project. URL: http://www.itee.uq.edu.au/~eresearch/projects/dart/
[100] DART Metadata Schema Registry. URL: http://www.itee.uq.edu.au/~eresearch/projects/dart/outcomes/metadataschemareg.html
[101] Institutional Profiling and Terms & Conditions Services Scoping Study, April 2004. URL: http://edina.ac.uk/projects/iptc/InstitutionalProfilingFinalReport-execsum.pdf
[102] Moving forward with shared services: shared infrastructure for the JISC Integrated Information Environment. Annex A JIIE(05)36
[103] IESR Metadata Review September 2005: Background Information. URL: http://iesr.ac.uk/metadata/reviews/mdrevbkd.html
[104] GeoCrossWalk. URL: http://www.geoxwalk.ac.uk/
[105] HILT. URL: http://hilt.cdlr.strath.ac.uk/
[106] Heery, Rachel. *Delivering HILT as a JISC IE shared service*. October 2003. URL: http://www.ukoln.ac.uk/metadata/hilt/m2m-report/hilt-final-report.pdf
[107] HILT M2M Demonstrator Feasibility Study. Final Report, March 2005. URL: http://hilt.cdlr.strath.ac.uk/hiltm2mfs/index.html
[108] Becta Vocabulary Studio. URL: http://www.becta.org.uk/studio/
[109] Connotea. URL: http://www.connotea.org/
[110]Indexing and Authority Files, an Overview. URL: http://www.paradigm.ac.uk/workbook/cataloguing/indexing.html
[111] In Between weblog. URL: http://digilib.weblog.ub.rug.nl/archive/2006/5/19
[112] Lorcan Dempsey's weblog. URL: http://orweblog.oclc.org/archives/001022.html
[113] JISC Information Environment and Digital Repositories Workshop, 22 May 2006. URL: http://www.jisc.ac.uk/rep_www06_pres.html
[114] OCLC name lookup web service. URL: http://alcme.oclc.org/eprintsUK/index.html
[115] In Between weblog. URL: http://digilib.weblog.ub.rug.nl/archive/2006/5/19
[116] *International Standard for Archival Authority Records (Corporate, Personal and Family) ISAAR(CPF).* URL: http://www.icacds.org.uk/eng/standards.htm
[117] Access to Archives. URL: http://www.a2a.org.uk/
[118] Oxford Dictionary of National Biography. URL: http://www.oxforddnb.com/
[119] STLQ. URL: http://stlq.info/2006/05/scopus_author_identifier_new_f.html
[120] IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). URL: http://www.ifla.org/VII/d4/wg-franar.htm
[121] Outline for ISO standard ISPI (International Standard Party Identifier Code). URL: http://www.collectionscanada.ca/iso/tc46sc9/docs/sc9n429.pdf
[122] Atkinson, M. et al. *Computer Challenges in e-Science*. URL: www.semanticgrid.org/docs/Vision.pdf
[123] JISC briefing paper on e-Infrastructure. May 2006. URL: http://www.jisc.ac.uk/programme_einfrastructure.html
[124] Research Councils UK (RCUK). URL: http://www.rcuk.ac.uk/
[125] Office of Science and Innovation. URL: http://www.dti.gov.uk/science/index.html
[126] Research Councils UK (RCUK). Statement on open access to research results. 28 June 2006. URL: http://www.rcuk.ac.uk/20060628openaccess.asp
[127] Research Information Network. URL: http://www.rin.ac.uk/
[128] National Health Service (NHS). URL: http://www.nhs.uk/
[129] Strategic e-Content Alliance (SEA). URL: http://www.jisc.ac.uk/sea.html
[130] The British Library (BL). URL: http://www.bl.uk/
[131] The British Broadcasting Corporation (BBC). URL: http://www.bbc.co.uk/
[132] British Education Communication and Technologies Agency (BECTa). URL: http://www.becta.org.uk/
[133] UK e-Science Core Programme. URL: http://www.epsrc.ac.uk/ResearchFunding/Programmes/e-Science/default.htm

[134] MLA. URL: http://www.mla.gov.uk/
[135] NHS National Library for Health (NLH). URL: http://www.library.nhs.uk/Default.aspx
[136] Common Information Environment (CIE). URL: http://www.common-info.org.uk/
[137] JISC PALS Metadata and Interoperability Working Group. URL: http://www.jisc.ac.uk/index.cfm?name=programme_pals2
[138] Access Management Federation. URL: http://www.jisc.ac.uk/federation.html
[139] McKenna, R. JISC (draft) Software Quality Assurance Policy. 27 Aug. 2004, p4
[140] SourceForge. URL: http://sourceforge.net/
[141] Gforge. URL: http://gforgegroup.com/
[142] Dublin Core Metadata Initiative. URL: http://uk.dublincore.org/
[143] DCMI Affiliate Programme: UKOLN activities. URL: http://www.ukoln.ac.uk/metadata/dcmi/